Depression-Free Days as a Summary Measure of the Temporal Pattern of Response and Remission in the Treatment of Major Depression: A Comparison of Venlafaxine, Selective Serotonin Reuptake Inhibitors, and Placebo

Rajiv Mallick, Ph.D.; Jieling Chen, M.A.;
A. Richard Entsuah, Ph.D.; and Alan F. Schatzberg, M.D.

Background: This article develops and applies depression-free days (DFDs) as a summary measure of the temporal pattern of response and remission in a comparison of venlafaxine (a dual-action serotonin-norepinephrine reuptake inhibitor) with selective serotonin reuptake inhibitors (SSRIs) and placebo.

Method: Weekly data on the 17-item Hamilton Rating Scale for Depression (HAM-D-17) from 2046 patients with DSM-III-R/IV–established moderate-to-severe major depression, participating in 1 of 8 randomized, double-blind, controlled studies that compared venlafaxine with an SSRI (fluoxetine, paroxetine, or fluvoxamine) or with both placebo and an SSRI, were used to estimate DFDs. Maximum DFDs were imputed to maintained HAM-D-17 scores ≤ 7 (asymptomatic depression) over time, minimum DFDs to persistent HAM-D-17 scores ≥ 15 (acutely symptomatic depression), and prorated DFDs to intermediate HAM-D-17 scores. A secondary construct was developed to test sensitivity to a less stringent upper threshold of acutely symptomatic depression (HAM-D-17 score ≥ 22). Using a tertiary construct, sensitivity to a more stringent lower threshold representing elimination of residual symptoms was also evaluated. The construct validity of the primary and the secondary DFDs measures was assessed in terms of their correlation with sustained low clinical global severity of illness (scores of 1 or 2 on the Clinical Global Impressions-Severity of Illness scale). For each construct, DFDs were compared across the 3 treatment groups and corresponding effect sizes were generated.

Results: Overall, sustained low clinical global severity of illness was associated with 38.3 median (interquartile range, 29.8 to 44.2) DFDs relative to 5.7 (interquartile range, 0 to 20.6) median DFDs associated with nonsustained low clinical global severity; similar differences emerged in terms of sustained asymptomatic depression. The venlafaxine group (N = 851) experienced a median of 18.8 (interquartile range, 0.4 to 34.6) DFDs compared with a median of 13.6 (interquartile range, 0 to 29.8) DFDs in the SSRI group (N = 749) and 7.4 (interquartile range, 0 to 26.2) DFDs in the placebo group (N = 446) ($p < .0001$ overall; venlafaxine vs. SSRIs, $p = .0015$, effect size = 0.2; venlafaxine vs. placebo, $p < .0001$, effect size = 0.4; and SSRIs vs. placebo, $p = .0007$, effect size = 0.2). The secondary and tertiary DFDs constructs yielded similar, albeit narrower, differences in all comparisons.

Conclusion: The construct of DFDs was found to be a useful summary measure of sustained remission. Active treatments were associated with more DFDs than placebo, and venlafaxine with more DFDs than SSRIs, consistent with corresponding differences in sustained remission.

(J Clin Psychiatry 2003;64:321–330)

Received August 22, 2001; accepted August 15, 2002. From the Department of Psychiatry and Behavioral Sciences, School of Medicine, Stanford University, Stanford, Calif. (Dr. Schatzberg); and Global Health Outcomes Assessment (Dr. Mallick and Ms. Chen) and the Department of Biostatistics (Dr. Entsuah), Wyeth Research, Collegeville, Pa.

Preparation of this manuscript was supported by an unrestricted educational grant from Wyeth, Collegeville, Pa.

Dr. Schatzberg has served as a consultant for and has received honoraria from Wyeth, Lilly, Glaxo, Solvay, Organon, Merck, and Corcept Therapeutics; has received grant/research support from Wyeth and Glaxo; has served on the speakers/advisory boards for Wyeth and Lilly; and is a major stock shareholder for Merck, Corcept Therapeutics, and Pfizer.

Corresponding author and reprints: Alan F. Schatzberg, M.D., Chairman, Department of Psychiatry and Behavioral Sciences, School of Medicine, Stanford University, Stanford, CA 94305-5717.

In clinical trials, antidepressants are typically evaluated in terms of efficacy measures (e.g., the Hamilton Rating Scale for Depression [HAM-D],[1] the Montgomery-Asberg Depression Rating Scale [MADRS],[2] or the Clinical Global Impressions-Severity of Illness [CGI-S] or -Improvement [CGI-I] scales).[3] Moreover, efficacy is usually assessed in terms of average reduction in HAM-D scores or attainment of a minimum response, typically a 50% decline in HAM-D or MADRS scores between baseline and end of treatment.[4,5] Both epidemiologic[6] and clinical studies[7] have demonstrated, however, that the 50% improvement norm may result in continued depressive symptoms for many patients.

Remission from acute depressive illness (i.e., minimal residual clinical symptoms) is less often examined as an

explicit outcome in clinical trials, and sustained remission and prevention of relapse are rarely addressed. In studies that do examine remission, it has been observed that of the patients who respond during acute treatment, only about half subsequently progress to remission.[8] In addition, since major depression is characterized by recurrent episodes as well as relapses within episodes,[9,10] the existence of residual symptoms during acute response to treatment is also likely to be associated with a full relapse or recurrence.[11–13]

Less than optimal remission of major depressive symptoms is known to have adverse health outcomes implications. First, absence of full remission is associated with greater use of health care services—visits to general practitioners or psychiatric specialists for reevaluation and for changes in prescribed medication,[14] emergency room visits, and hospitalization.[15] Second, patients who do not achieve full remission or subsequently relapse are likely to have greater risk of disability and absenteeism.[16–19] Finally, relapse or recurrence of symptomatic major clinical depression has been found to be associated with substantially impaired quality of life for patients,[20–22] including greater risk of suicide.[16]

In light of the above, the importance of remission is increasingly recognized.[23,24] It is also understood that remission and relapse in major depressive symptoms must be operationally defined if these concepts are to be meaningfully applied in clinical research. Thus, attainment of a HAM-D score ≤ 7 has become generally accepted as evidence of asymptomatic depression, i.e., "remission," and a HAM-D score ≥ 15 as evidence of being fully symptomatic.[25] Further, a return from remission to a fully symptomatic state within the same depressive episode is characterized as "relapse" and the occurrence of a new episode as "recurrence."[25]

It is also recognized that a static conceptualization of the attainment of remission or the occurrence of relapse may be insufficient and that dynamic considerations that incorporate duration in each of the clinical states may be useful.[26] First, attainment of response typically leads to a state of "partial remission" (defined as HAM-D-17 in the range between, but not including, 7 and 15), from which state a patient could either move on to full remission, relapse, or remain in the state of partial remission or "flurry," i.e., a relatively short period outside the asymptomatic range that does not constitute relapse into a fully symptomatic state.[25] Second, remission is required to be a minimum of 2 weeks in duration, and if it is not maintained for longer periods, it is characterized as "transient."[27]

This article is motivated by the belief that despite a recognition of the importance of remission[23,24] and its duration,[28] the relative complexity of fully characterizing the time pattern of response or remission has prevented the use of temporal measures of treatment success or failure in clinical trials of major depression. On the other hand, measuring response or even remission rates at a single point in time, while simpler, is likely to be insensitive to the remitting-relapsing course of depressive symptoms. Accordingly, a valid, yet simple and clinically meaningful quantitative assessment that integrates occurrence and duration of time in remission (sustained or transient), response without remission, and nonresponse during a major depressive episode is needed. Such a measure would be potentially useful in discriminating between the effects of being in one or other of these states as well as in describing effectiveness of treatments.

In this article, we propose depression-free days (DFDs), previously used in comparing psychotherapy regimens with pharmacologic treatment,[29] as a summary measure of the temporal pattern of response and remission that can be flexibly constructed from sequential assessments of the HAM-D-17 during treatment of a major depressive episode. Using pooled data from a previously published meta-analysis that compared venlafaxine (XR; a dual serotonin-norepinephrine reuptake inhibitor), selective serotonin reuptake inhibitors (SSRIs), and placebo,[30] we develop a primary DFDs measure based on the well-accepted thresholds of HAM-D-17 score ≤ 7 (remission or asymptomatic depression) and HAM-D-17 score ≥ 15 (fully symptomatic depression),[25] and prorating of time in remission or symptomatic depression according to proximity of HAM-D-17 scores to these specific anchors. By this construct, DFDs are optimized for patients who achieve and sustain remission, especially those with longer sustained remission. In contrast, patients who never respond, those who respond but never attain remission, and those who attain remission but subsequently relapse are expected to have, on average, fewer DFDs.

We also develop a secondary construct based on a less conservative definition of what constitutes a fully symptomatic state (HAM-D-17 score ≥ 22),[29] which attributes relatively more DFDs even when there is response rather than remission. Finally, we explore a tertiary construct based on a more conservative definition of treatment success in terms of elimination of residual symptoms (HAM-D-17 score = 6, 5, 4, 3, 2, 1), which sets a higher standard for optimal depression-free time than attainment of a HAM-D score of 7. We evaluate the validity, interpretability, and efficiency of DFDs in variably discriminating between various patient health states as well as across treatment groups for each of these constructs.

METHOD

Design of the Clinical Studies

In these studies (references 31–36 as well as data on file, Wyeth Laboratories), previously summarized,[30] conducted in the United States, Europe, or Canada, patients were randomly assigned to treatment with venlafaxine

  © Copyright 2003 Physicians Postgraduate Press, Inc.

(75–375 mg/day immediate release or 75–225 mg/day XR), an SSRI (fluoxetine [20–80 mg/day], paroxetine [20–40 mg/day], or fluvoxamine [100–200 mg/day]), or placebo (4 studies only) during the double-blind treatment period. Seven studies were outpatient trials; the eighth study[32] enrolled only inpatients. Four studies were of 8-week duration, 3 studies were of 6-week duration, and 1 study was of 12-week duration[30]; for consistency, pooled data were analyzed at 8 weeks with data truncated at 8 weeks for the 12-week study and extrapolated to 8 weeks for the 6-week studies. In sensitivity analysis, we also examined the implications, however, of truncating the data at 6 weeks for all studies. The studies were conducted with institutional review board approval at participating centers and according to the guidelines of the Declaration of Helsinki and its amendments, with all patients providing explicit written informed consent.

Patient Population

Eligibility for all studies included in this analysis required patients to be at least 18 years old and to meet the criteria of the *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition, Revised (DSM-III-R)[37] or Fourth Edition (DSM-IV)[38] for major depression or major depressive disorder, respectively, for at least 1 month before enrollment. Further, eligible patients were typically required to have minimum scores (depending on the study) of either 20 on the HAM-D-21 or 25 on the MADRS[2] at baseline.

Patients with clinically significant physical or mental comorbidities (e.g., cardiovascular disease, renal or hepatic disorders, seizure disorders, or recent history of alcohol or drug abuse) were excluded from study participation. Patients who had known hypersensitivity to the study drugs or who had used investigational or antipsychotic drugs within 30 days, a monoamine oxidase inhibitor within 14 days, or other antidepressant, anxiolytic, or sedative-hypnotic drug within 7 days of the double-blind treatment period were also excluded. Patients who had taken any nonpsychopharmacologic drugs with psychotropic effects within 7 days of the double-blind treatment period were excluded unless a stable dosage of the drug had been maintained for at least 1 month prior to study start. The sociodemographic and pretreatment clinical characteristics of the pooled study groups have been previously reported.[30]

Efficacy and Safety Assessments in Individual Studies

For each of the 8 studies pooled for this analysis, evaluations of the HAM-D-17 total score and the CGI-S score were performed at baseline, prior to double-blind therapy, and subsequently on study days 7, 14, 21, 28, 42, and, if available, 56. In the individual studies, response was defined in terms of a 50% reduction from the baseline total scores for the HAM-D-17. Safety and tolerability were evaluated on the basis of reported adverse events throughout the study evaluation period and also on the basis of any changes that occurred in the physical examination, vital signs, 12-lead electrocardiogram recordings, or clinical laboratory tests during treatment. Because this report focuses on effectiveness in terms of depression-free days derived from the clinical efficacy measures, safety comparisons are restricted to only the proportions of patients withdrawn from double-blind therapy due to side effects or lack of efficacy.

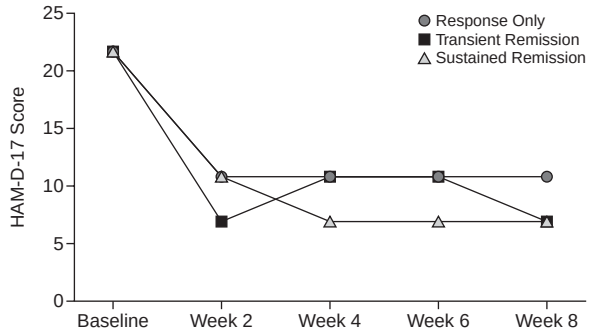
Concepts, Definitions, and Outcomes

In order to evaluate its relationship to DFDs, remission was defined broadly in terms of 2 alternative measures: (1) low clinical global severity (CGI-S) score of 1 ("normal, not at all ill") or 2 ("borderline mentally ill") and (2) asymptomatic depression (HAM-D-17 score ≤ 7), following precedent using similar approaches.[39] We required patients with sustained remission to attain a state of low clinical global severity or asymptomatic depression continuously for at least 4 weeks; by definition, this was incompatible with any evidence of relapse or flurry from the point of attainment of remission.[39] In contrast, achieving remission for less than 4 weeks or being followed by periods of relapse or partial remission (flurry) was defined as transient remission, following similar approaches.[27,39]

These concepts were operationalized by estimating sustained remission rates as the proportion of patients who had achieved remission by week 2 or week 4 and subsequently maintained remission through the end of the 8-week analysis period, with no evidence of relapse or flurry, alternatively in terms of the CGI-S or the HAM-D-17 criteria. We further classified patients who achieved sustained remission by the duration of time they were in continuous remission, i.e., as 4-week or 6-week sustained remitters. Patients who did not achieve sustained remission were also subclassified, albeit somewhat differently, by the CGI-S criteria and the HAM-D-17 criteria. On the basis of the former, the nonsustained remitters were classified as (1) nonremitters: those who never attained a CGI-S score of 1 or 2; or (2) transient remitters: those who attained a CGI-S score of 1 or 2 but for less than 4 continuous weeks. On the basis of the latter, the nonsustained remitters were classified as (1) nonresponders: those who did not respond (i.e., never attained a 50% reduction in HAM-D-17 scores from baseline); (2) responders only: those who responded but never attained a HAM-D-17 score ≤ 7; or (3) transient remitters: those who attained a HAM-D-17 score ≤ 7 at some point during the study period but subsequently slipped into partial remission or full relapse, and were never observed to be in continuous remission for more than 3 weeks.

Depression-free days were estimated, for each patient, using weekly scores on the HAM-D-17. For the primary

Figure 1. Time Pattern of 17-Item Hamilton Rating Scale for Depression (HAM-D-17) Scores and Associated Depression-Free Days (DFDs)[a]



[a]Median DFDs for each response group were as follows: response only, 14.0 (30.8), transient remission, 28.9 (37.8), sustained remission, 38.5 (45.2). Values in parentheses denote DFDs for the secondary construct.

measure of DFDs, a value of 7 DFDs was assigned when a HAM-D-17 score ≤ 7 (the remission threshold) was maintained over a weekly interval, a value of 0 DFDs when a HAM-D-17 score ≥ 15 (acutely symptomatic depression) persisted over the week, and weighting DFDs proportionately to HAM-D-17 scores that varied between 7 and 15, i.e., in the partial remission range, during the week. For the secondary DFDs construct, the above was modified to attribute 0 DFDs only when patients had a persistent HAM-D-17 score ≥ 22 (rather than 15) and weigh DFDs proportionately to HAM-D-17 scores between 7 and 22, thus imputing some DFDs to scores between 15 and 22. For the tertiary construct, it was the lower threshold that was modified to reflect a more stringent definition of remission in terms of elimination of residual symptoms, i.e., HAM-D-17 score = 1. The tertiary construct is reported on only for treatment-related analyses for brevity. For all 3 versions, weekly DFDs were aggregated over the 8-week analysis period, and therefore by construct, DFDs could potentially vary between 0 and 56 days, as described in Appendix 1. Using 3 hypothetical temporal patterns of HAM-D-17 scores representing response only (without remission), transient remission (remitting-relapsing course of symptoms), and sustained remission, Figure 1 presents calculated DFDs for each longitudinal pattern (with DFDs for the secondary construct in parentheses). It shows (1) that the DFDs measure, whether the primary or secondary construct is used, distinguishes these 3 states while a point-in-time assessment of remission at week 8, for example, would fail to distinguish sustained remission from transient remission and (2) that the primary construct of DFDs distinguishes the "response only" time pattern better than the secondary construct since the latter also attributes some DFDs to HAM-D scores between 15 and 22.

Validation of the DFDs Construct

Using combined data over all treatment groups, we assessed the evaluative properties[40] of the DFDs measure, i.e., whether the construct enables as its primary goal a useful and accurate distinction between patients with and without sustained remission. We examined the construct validity[41] of the DFDs measure primarily in terms of its association with an external, independent measure: sustained low clinical global severity of illness (CGI-S score of 1 or 2). In addition, we evaluated the relationship of DFDs to sustained asymptomatic depression (HAM-D-17 score ≤ 7). Both criteria required sustained remission for at least 4 weeks, i.e., that there be no evidence of relapse or flurry after remission is attained. For each test, we compared median DFDs between patients who achieved sustained remission and those who did not, using the rank-order Wilcoxon 2-sample test. We further evaluated the validity of the DFDs measure in terms of its intended secondary goals to (1) distinguish patients classified according to varying durations of continuous remission (4 weeks or 6 weeks) and (2) distinguish patients with transient remission from patients with response but no remission and those with no response, using the rank-order Kruskal-Wallis multiple-sample test. Each of these exercises was replicated for the secondary DFDs construct.

Treatment-Related Analyses and Statistical Issues

The analyses were performed on a modified intent-to-treat (ITT) sample, which included all patients who received at least 1 dose of study medication and had at least 1 evaluation during therapy. All outcomes were estimated using the last-observation-carried-forward method, consistent with previous literature on DFDs[29] and remission,[30] which allowed the inclusion of patients who were withdrawn early. To evaluate if this approach caused potential bias in treatment group–related analyses, we compared the proportions of patients thus affected, i.e., those who withdrew early, across treatment groups.

For treatment-related analyses, analysis of variance was used first to assess the comparability of the 3 treatment groups (venlafaxine, SSRIs, or placebo) with respect to age, weight, duration of depressive episode, and baseline HAM-D-17 and MADRS total scores. Baseline severity of depression was also assessed using the CGI-S item. In addition, DFDs were compared across treatment groups after first testing the underlying distribution for normality. Because standard normality assumptions were found not to hold, nonparametric methods (rank-order Kruskal-Wallis tests) were used to compare the distribution of DFDs across treatment groups. Sensitivity of treatment group–related differences in DFDs was examined in terms of the alternative construct of fully asymptomatic depression (HAM-D-17 score ≥ 22) as well as in

Table 1. Association of Depression-Free Days (DFDs) With Remission and Sustained Remission, Defined in Terms of Low Clinical Global Impressions-Severity of Illness Scale (CGI-S) Scores of 1 or 2[a,b,c]

Outcome in Terms of Low Clinical Global Severity (CGI-S score = 1 or 2)	Patients (total N = 2046) N	%	DFDs, Median (interquartile range) Primary Construct	Secondary Construct
(i) Never attained low clinical global severity	1028	50.2	0 (0 to 8.3)	13.0 (1.9 to 24.0)
(ii) Transient low clinical global severity (for 3 or fewer weeks)	516	25.2	22.3 (12.9 to 31.9)	32.2 (25.7 to 39.2)
(iii) Nonsustained low clinical global severity (group i + group ii)	1544	75.5	5.7 (0 to 20.6)	21.5 (7.0 to 32.0)
(iv) Sustained low clinical global severity for 4 to 5 weeks	343	16.8	35.0 (28.4 to 39.8)	41.1 (35.2 to 45.3)
(v) Sustained low clinical global severity for ≥ 6 weeks	159	7.8	45.5 (40.3 to 48.6)	48.8 (45.3 to 51.3)
(vi) All sustained low clinical global severity (group iv + group v)	502	24.5	38.3 (29.8 to 44.2)	43.9 (37.3 to 47.8)

[a]Calculations based on data also separately analyzed in Thase et al.[30]
[b]$p < .0001$, Kruskal-Wallis rank-order test of differences in medians across groups i, ii, iv, and v; $p < .0001$, Wilcoxon rank-order test of differences in medians between groups vi and iii.
[c]Effect size (group vi vs. group iii) = (median group vi – median group iii)/pooled interquartile range. For primary construct, effect size = 1.02; for secondary construct, effect size = 0.81. Tertiary construct not presented here for brevity.

terms of successive reductions in the lower threshold to reflect not only remission but virtual elimination of residual depressive symptoms (i.e., HAM-D-17 score = 6, 5, 4, 3, 2, 1).

Interpretability of DFDs

In addition to validity, interpretability of a new measure in terms of other clinical criteria is important.[42] Following similar approaches,[43] we sought to evaluate prospectively the interpretability of potential treatment-related differences in DFDs in terms of implications for the effect size.[44,45] The effect size approximates the mean difference between treatment groups deflated by the pooled standard deviation,[44] or, as in this article, the ratio of the difference in median DFDs across treatment groups to the pooled interquartile range of DFDs, when nonparametric comparisons are more appropriate.[46] Effect sizes over 0.2 are considered clinically meaningful, those within the range of 0.5 to 0.8 are considered moderately so, and those over 0.8 are considered substantial.[44,47,48]

RESULTS

Baseline Characteristics of Patients

In total, 2117 patients were enrolled in the 8 trials; 2046 (96.6%) were included in the modified ITT analyses of the effects of venlafaxine and venlafaxine XR (N = 851), the SSRIs (N = 749), and placebo (N = 446). Approximately two thirds of the patients were women, and all treatment populations had similar characteristics at baseline. However, patients enrolled in the placebo-controlled studies were less severely depressed than those enrolled in the studies that included only an active comparator. Significantly fewer patients randomly assigned to placebo had CGI-S scores of > 4 at baseline ($p < .001$) when compared with subjects randomly assigned to receive venlafaxine or the SSRIs (36%, 53%, and 53%, respectively). Across the placebo-controlled studies, there was no difference between treatment groups in baseline CGI-S scores.

Patient Withdrawal

In total, 83 patients (9%) assigned to venlafaxine were withdrawn from treatment due to side effects, compared with 57 patients (7%) assigned to SSRI and 10 patients (2%) assigned to placebo (Fisher exact test; venlafaxine vs. placebo and SSRI vs. placebo [$p = .001$]; venlafaxine vs. SSRI [$p = .185$] comparison was not significant).[30] Of the 895 patients assigned to venlafaxine, 33 (4%) were withdrawn due to lack of efficacy, compared with 46 (6%) of the 769 patients assigned to an SSRI and 63 (14%) of the 453 patients assigned to placebo (Fisher exact test; venlafaxine vs. SSRI [$p = .037$]; venlafaxine vs. placebo [$p = .001$]; SSRI vs. placebo [$p = .001$]).[30] Thus, overall discontinuation rates were approximately similar: 13% in each of the venlafaxine and SSRI groups and 16% in the placebo group.

Validation of the DFDs Construct: Pooled Data Analysis

Prior to analyzing treatment-related DFDs, data pooled over all treatment groups were used to compare DFDs by whether or not patients were sustained remitters. Table 1 reports this relationship in terms of the independent measure of low clinical global severity (CGI-S score). Patients with sustained low CGI-S scores (of 1 or 2) for at least 4 weeks had an estimated median of 38.3 (interquartile range, 29.8 to 44.2) DFDs compared with 5.7 (interquartile range, 0 to 20.6) DFDs for patients who did not achieve the sustained low CGI-S criteria ($p < .0001$), a difference effect size of 1.02 ([38.3 less 5.7] divided by pooled interquartile range [31.9] of all patients). Also, DFDs varied significantly with the duration of sustained low CGI-S scores (Table 1). A similar association emerged between the secondary DFDs construct (attributing 0 DFDs for HAM-D-17 scores ≥ 22) and both the attainment and duration of sustained low CGI-S scores ($p < .0001$) (Table 1), albeit with somewhat smaller numerical differences across categories and effect size of differences compared with the primary construct.

Table 2. Association of Depression-Free Days With Response, Remission, and Sustained Remission, Defined in Terms of Asymptomatic Depression (HAM-D-17 score ≤ 7)[a,b,c]

Outcomes in Terms of Asymptomatic Depression (HAM-D-17 score ≤ 7)	Patients (total N = 2046) N	%	DFDs, Median (interquartile range) Primary Construct	Secondary Construct
(i) No response (< 50% reduction in baseline HAM-D-17 score)	723	35.3	0 (0 to 0)	6.5 (0 to 16.8)
(ii) Response but not asymptomatic depression (never attained HAM-D-17 score of 7)	444	21.7	11.4 (5.7 to 19.7)	25.9 (19.4 to 32.0)
(iii) Transient asymptomatic depression (HAM-D-17 score ≤ 7 for 3 or fewer weeks)	507	24.8	26.7 (18.8 to 32.8)	35.2 (29.4 to 40.6)
(iv) Nonsustained asymptomatic depression (group i + group ii + group iii)	1674	81.8	7.9 (0 to 22.8)	22.9 (8.4 to 33.1)
(v) Sustained asymptomatic depression (HAM-D-17 score ≤ 7 for 4 to 5 weeks)	275	13.4	39.4 (35.9 to 42.4)	44.6 (41.5 to 47.1)
(vi) Sustained asymptomatic depression (HAM-D-17 score ≤ 7 for ≥ 6 weeks)	97	4.7	47.7 (45.9 to 49.0)	50.4 (48.8 to 52.0)
(vii) All sustained asymptomatic depression (group v + group vi)	372	18.2	41.6 (37.6 to 46.4)	46.2 (42.6 to 49.7)

[a]Calculations based on data also separately analyzed in Thase et al.[30]
[b]$p < .0001$, Kruskal-Wallis rank-order test of differences in medians across groups i, ii, iii, v, and vi; $p < .0001$, Wilcoxon rank-order test of differences in medians between groups vii and iv.
[c]Effect size (group vii vs. group iv) = (median group vii – median group iv)/pooled interquartile range. For primary construct, effect size = 1.06; for secondary construct, effect size = 0.86. Tertiary construct not presented here for brevity.
Abbreviation: HAM-D-17 = 17-item Hamilton Rating Scale for Depression.

Table 3. Depression-Free Days (DFDs) Across Treatment Groups[a]

Efficacy Measure[b]	Venlafaxine (N = 851)	SSRIs (N = 749)	Placebo (N = 446)	p Values
Primary construct[c]	18.8 (0.4 to 34.6)	13.6 (0 to 29.8)	7.4 (0 to 26.2)	< .0001[ϕ]
				.0015*
				< .0001[#]
				.0007[ρ]
Secondary construct[d]	29.9 (15.2 to 41.3)	26.8 (11.9 to 38.7)	23.2 (8.4 to 36.2)	< .0001[ϕ]
				.0046*
				< .0001[#]
				.0196[ρ]
Tertiary construct[e]	11.3 (0.3 to 23.3)	7.8 (0 to 18.5)	4.3 (0 to 15.8)	< .0001[ϕ]
				.0008*
				< .0001[#]
				.0006[ρ]

[a]Calculations based on data also separately analyzed in Thase et al.[30]
[b]DFDs, median (interquartile range) for each construct.
[c]Primary DFDs construct based on attributing 1 DFD to HAM-D-17 score ≤ 7, 0 DFDs to HAM-D-17 score ≥ 15, and prorating intermediate values.
[d]Secondary DFDs construct based on attributing 1 DFD to HAM-D-17 score ≤ 7, 0 DFDs to HAM-D-17 score ≥ 22, and prorating intermediate values.
[e]Tertiary (most stringent) DFDs construct based on attributing 1 DFD to HAM-D-17 score ≤ 1, 0 DFDs to HAM-D-17 score ≥ 15, and prorating intermediate values.
[ϕ]Overall differences across the 3 treatment groups (Kruskal-Wallis rank-order, multiple-sample test).
*Venlafaxine compared with SSRIs (Wilcoxon rank-order test).
[#]Venlafaxine compared with placebo (Wilcoxon rank-order test).
[ρ]SSRIs compared with placebo (Wilcoxon rank-order test).
Abbreviations: HAM-D-17 = 17-item Hamilton Rating Scale for Depression, SSRI = selective serotonin reuptake inhibitor.

Also, DFDs were similarly associated with achievement and duration of sustained remission when defined alternatively in terms of asymptomatic depression (HAM-D-17 score ≤ 7 for at least 4 weeks) ($p < .0001$). This relationship held in terms of the primary DFDs construct as well as, albeit with narrower numerical differences, for the secondary DFDs construct (Table 2).

Depression-Free Days: Differences by Treatment Group

Over the average 8-week period, patients in the venlafaxine group (N = 851) had an estimated median of 18.8 DFDs compared with a median of 13.6 DFDs in the SSRI group (N = 749) and a median of 7.4 DFDs in the placebo group (N = 446), based on the primary construct (Table 3). Mean DFDs varied in the same direction across treatment groups. The treatment-related differences in DFDs translated into an effect size of 0.2 ([18.8–13.6]/31.9) between the venlafaxine group and the SSRI groups, an effect size of 0.4 between the venlafaxine group and the placebo group, and an effect size of 0.2 between the SSRI group and the placebo group.

In terms of the secondary construct, DFDs were uniformly greater for all treatment groups, and differences between the treatment groups, while numerically smaller, retained their statistical significance (Table 3). Using the tertiary construct, DFDs were uniformly lower but differences across treatment groups were still significant, albeit similar in numerical magnitude to the secondary construct (Table 3). Finally, DFDs were found to vary significantly in the same direction across treatment groups when estimated by truncating all pooled data at 6 rather than 8 weeks.

There were similarly significant treatment-related differences in the distribution of patients by achievement and duration of sustained low clinical global severity of illness (CGI-S score = 1 or 2) (Table 4) or asymptomatic depression (HAM-D-17 score ≤ 7) (Table 5). Compared with placebo, both the active treatment groups were associated with a greater proportion of patients who achieved sustained low clinical global severity or asymptomatic depression; active treatments were also associated with longer duration in these states. Venlafaxine was associated with a greater proportion of patients with sustained

Table 4. Distribution of Patients by Remission and Duration of Remission: Sustained Low Clinical Global Severity Criteria (CGI-S score = 1 or 2)[a]

Outcomes in Terms of Low Clinical Global Severity (CGI-S score = 1 or 2)	Venlafaxine (N = 851)		SSRIs (N = 749)		Placebo (N = 446)	
	N	%	N	%	N	%
Never attained low clinical global severity	378	44.4	382	51.0	268	60.1
Transient (for 3 or fewer weeks of continuous) low clinical global severity	244	28.7	178	23.8	94	21.1
4 to 5 weeks of continuous low clinical global severity	143	16.8	140	18.7	60	13.5
≥ 6 weeks of continuous low clinical global severity	86	10.1	49	6.5	24	5.4

[a]Calculations based on data also separately analyzed in Thase et al.[30] p Values based on chi-square test of differences in frequency distribution across treatment groups: overall differences in distribution across the 3 treatment groups, p = .001; venlafaxine vs. SSRIs, p = .003; venlafaxine vs. placebo, p = .001; SSRIs vs. placebo, p = .016.
Abbreviations: CGI-S = Clinical Global Impressions-Severity of Illness scale, SSRI = selective serotonin reuptake inhibitor.

Table 5. Distribution of Patients by Response, Remission, and Duration of Remission: Sustained Asymptomatic Depression Criteria (HAM-D-17 score ≤ 7)[a]

Outcomes in Terms of Asymptomatic Depression (HAM-D-17 score ≤ 7)	Venlafaxine (N = 851)		SSRIs (N = 749)		Placebo (N = 446)	
	N	%	N	%	N	%
No response (< 50% reduction in baseline HAM-D-17 score)	255	30.0	267	35.6	201	45.1
Response but no asymptomatic depression	162	19.0	185	24.7	97	21.7
Transient (for 3 or fewer weeks of continuous) asymptomatic depression	238	28.0	168	22.4	101	22.6
Sustained asymptomatic depression for 4 to 5 weeks	140	16.5	101	13.5	34	7.6
Sustained asymptomatic depression for ≥ 6 weeks	56	6.6	28	3.7	13	2.9

[a]Calculations based on data also separately analyzed in Thase et al.[30] p Values based on chi-square test of differences in frequency distribution across treatment groups: overall differences in distribution across the 3 treatment groups, p = .001; venlafaxine vs. SSRIs, p = .001; venlafaxine vs. placebo, p = .001; SSRIs vs. placebo, p = .002.
Abbreviations: HAM-D-17 = 17-item Hamilton Rating Scale for Depression, SSRI = selective serotonin reuptake inhibitor.

low clinical global severity or asymptomatic depression compared with both SSRIs and placebo, especially when comparing patients who achieved and sustained either measure of remission for at least 6 weeks (Tables 4 and 5). The SSRI group was associated with a significantly greater proportion of patients with sustained asymptomatic depression and low clinical global severity of illness compared with the placebo group. At the other end of the spectrum, both active treatments when compared with placebo were associated with a smaller proportion who had no response. In terms of the intermediate categories of response only or transient remission, the 3 groups were essentially similar, but SSRIs were associated with a higher proportion of patients who achieved response only but no remission compared with patients in the venlafaxine or placebo groups.

DISCUSSION

Although there is increasing recognition that remission is important, integrating remission-based outcomes in clinical trials and analyses, especially in terms of their time dependence, encounters many challenges. On the one hand, an outcome measure that evaluates remission, but does so at 1 timepoint alone, while being relatively simple, may not adequately quantify the potential benefits of sustainability or longer duration of sustained remission. On the other hand, interpreting findings that report a complex vector of time paths of remission, relapse, and response rates, even when there is consensus on their clinical meaningfulness, can be daunting for the practicing clinician. A further complication occurs when there are subtle disagreements on the thresholds (or anchors) that define depressive health states, i.e., fully symptomatic depression on the one hand and remission on the other.[25,29] This study suggests that depression-free days, an approximate summary measure of the amount of time that patients are expected to spend, on average, in a state of minimal depression symptoms, offers the potential of being a relatively simple, yet valid, flexible, and interpretable measure that can sensitively discriminate between sustained remission (and its duration), transient remission (i.e., remitting-relapsing symptoms), response without remission, and no response. These properties are discussed further below.

The DFDs measure was found to substantially distinguish patients with sustained remission (HAM-D-17 score ≤ 7 or CGI-S score of 1 or 2) from others, with a difference of about 33 DFDs and a corresponding effect size of slightly over 1. Further, the DFDs measure was also found to discriminate among patients with varying durations of sustained remission, a distinction emphasized in previous literature,[25,28–30] and among those who achieved only transient remission, response without remission, or no response, validating its stated purpose of distinguishing among these outcomes.

We found that by virtue of a flexible application of the underlying anchors, the DFDs measure also differentiated in terms of varying emphasis on response, remission, or elimination of all residual symptoms. Thus, the secondary construct, since it attributed some DFDs also to patients with minimal responses from baseline, yielded more DFDs overall, but the increase was disproportionately greatest in the placebo group, followed by the SSRI group, and least in the venlafaxine group. This finding was consistent with similar differences across treatment groups in the proportion of patients who achieved only a response but did not achieve remission. Accordingly, while the secondary DFDs construct maintained an ability to distinguish between sustained remitters and others, albeit with narrower differences, it discriminated more effectively than the primary construct between patients who achieved a response but not remission (i.e., a > 50% decline from the baseline HAM-D-17 score, but a HAM-D-17 score > 7). The differential level of discrimination between outcomes using the primary and secondary DFDs constructs suggests that DFDs may be flexibly constructed with varying specifications of the HAM-D–based anchors and yet be highly robust in terms of accomplishing their stated purpose of distinguishing patients who achieve (longer) sustained remission from others. Finally, in the tertiary construct, while overall DFDs declined as expected when the lower HAM-D-17 threshold was successively modified to reflect increasingly stringent criteria for what constitutes a fully depression-free day, treatment group–related differences persisted even in the most extreme case requiring complete elimination of residual symptoms as the criterion for a fully depression-free day.

Interpretability of the differences in DFDs in various comparisons was assessed in terms of effect sizes. The effect sizes of differences in DFDs between patients who did or did not meet our criterion of sustained remission, by either definition, were found to be the greatest, about 1, indicating a substantial ability of DFDs to discriminate between these 2 groups. Effect sizes of differences in DFDs between patients in adjacent subcategories were more modest yet moderate, ranging from 0.25 in the comparison of patients with 6-week and 4-week remission to 0.47 in the comparison of patients with transient remission to those with response alone. Effect sizes for the secondary DFDs construct were substantial but somewhat smaller, indicating that attribution of DFDs to patients who achieve limited response at HAM-D-17 scores between 15 and 22 dilutes somewhat the distinction between those who achieve full remission and others.

Overall variation in DFDs by sustained remission and its duration enables understanding of the sources of differences in DFDs across treatment groups: 19 DFDs for venlafaxine, relative to 14 DFDs for SSRIs and 7 DFDs for placebo. These differences in DFDs were consistent with corresponding differences across treatment groups in the proportion of patients who attained sustained low clinical global severity of illness (CGI-S score of 1 or 2), sustained asymptomatic depression (HAM-D score ≤ 7), and the duration of sustained remission by either criterion. Further, the markedly higher DFDs among patients who achieved remission as early as week 2 and sustained it for at least 6 weeks enables also a better perspective on the issue of fast onset of response. Fast onset, if not matched by remission and, ideally, sustained remission, is neither clinically meaningful nor likely to contribute substantially to DFDs. Consequently, the faster response onset of venlafaxine noted in previous research[49] is unlikely to have accounted for all of the differences in DFDs across treatment groups observed in this study if in fact it reflected only transient remission or no remission at all. However, longer studies extending into the continuation phase are required to more fully clarify how sustained remission relates to DFDs and differences by treatment group.

It is arguable that the treatment-related differences attained statistical significance solely because of the pooling of studies. Pooling of studies is especially problematic if the underlying individual studies are heterogeneous in terms of treatment effects. However, previous analysis of these pooled studies revealed the opposite. Based on the Breslow-Day test for homogeneity of the odds ratio of achieving remission, treatment-related differences were found to be consistent in all 8 of the included studies ($p = .28$).[30] In this study, treatment-related differences in DFDs were also found to be robust to truncating the pooled data to 6 weeks to reflect the duration of the shortest studies. Finally, the possibility of exclusion bias has been previously examined. Twelve excluded studies that were arguably worthy of inclusion in qualitative analysis were observed to be quite different in terms of the patient populations they treated, the strength and duration of therapy administered, and definitions of outcomes; nevertheless, as previously reported, none of these studies found treatment-related differences between venlafaxine, SSRIs, and placebo that were discordant with the results reported here.[30]

A couple of important caveats of the interpretations attached to the analyses presented here must be noted: First, while the use of the term *depression-free days* might imply a measure that incorporates daily observation of depression symptoms, such is clearly not the case. As noted, the construct of depression-free days is based on weekly assessments of the physician-administered (17-item) HAM-D. Thus, it may not adequately capture intraweek variations in disease symptoms such as, for example, the daily diary in asthma[50] may potentially accomplish. Nevertheless, daily diary–based assessments in other conditions with mood- and anxiety-related effects have been found to be correlated with weekly HAM-D scores.[51,52] In any case, departures of daily (intraweek) depressive symptoms from the weekly assessment of the

  © Copyright 2003 Physicians Postgraduate Press, Inc.

HAM-D are likely to be random and, at minimum, not systematically linked to treatment groups. Second, while we used a last-observation-carried-forward approach to impute missing values, mixed modeling or other random-effects techniques[53] may also be applied to model the stochastic nature of the time course of symptoms. Nevertheless, the fact that overall dropout rates were similar across treatment groups would suggest that the approach used is unlikely to cause bias in the treatment-related comparisons.

In conclusion, DFDs appears to be an outcome measure that offers the potential to detect meaningful treatment differences. Further, the DFDs construct can be flexibly applied to take account of alternative anchors that define either fully symptomatic depression on the one hand or full remission on the other. In this analysis, the construct was shown to be robust to such a flexible application and the resultant differences in DFDs were in accord with prior expectations of what the primary version of the construct emphasizes. It is hoped that this exercise will motivate similar endeavors by both researchers and practitioners in an attempt to meaningfully understand and simply but effectively summarize treatment-related differences in response, remission, sustained remission, and duration of sustained remission, which otherwise would be complex and perhaps infeasible to combine as outcome measures.

Drug names: fluoxetine (Prozac and others), fluvoxamine (Luvox and others), paroxetine (Paxil), venlafaxine (Effexor).

REFERENCES

1. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62
2. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134:382–389
3. National Institute of Mental Health. Clinical global impressions. Psychopharmacol Bull 1985;21:839–843
4. Prien R, Carpenter L, Kupfer D. The definition and operational criteria for treatment outcome of major depressive disorder: a review of the current research literature. Arch Gen Psychiatry 1991;48:796–800
5. Depression Guideline Panel. Depression in Primary Care, vol 2. Rockville, Md: US Dept Health Human Services, Agency for Health Care Policy and Research; 1993. AHCPR publication 93-0551
6. Judd LL, Akiskal HS, Paulus M. The role and clinical significance of subsyndromal depressive symptoms (SSD) in unipolar major depressive disorder. J Affect Disord 1997;45:5–18
7. Maier W, Gansicke M, Weifenbach O. The relationship between major and subthreshold variants of unipolar depression. J Affect Disord 1997;45:41–51
8. Keller MB, Gelenberg AJ, Hirschfeld RMA, et al. The treatment of chronic depression, pt 2: a double-blind, randomized trial of sertraline and imipramine. J Clin Psychiatry 1998;59:598–607
9. Keller MB, Sharpiro R, Lavori P, et al. Relapse in major depressive disorder: analysis with life table. Arch Gen Psychiatry 1982;39:905–910
10. Keller MB. Introduction: advances in the management of chronic depressive and anxiety disorders. J Clin Psychiatry 1997;58(suppl 13):3–4
11. Paykel ES, Ramana R, Cooper Z, et al. Residual symptoms after partial remission: an important outcome in depression. Psychol Med 1995; 25:1171–1180
12. Dawson R, Lavori PW, Coryell WH, et al. Maintenance strategies for unipolar depression: an observational study of levels of treatment and recurrence. J Affect Disord 1998;49:31–44
13. Judd LL, Akiskal HS, Maser JD, et al. Major depressive disorder: a prospective study of residual subthreshold depressive symptoms as predictor of rapid relapse. J Affect Disord 1998;50:97–108
14. Jonsson B, Bebbington P. What price depression? the cost of depression and the cost-effectiveness of pharmacological treatment. Br J Psychiatry 1994;164:665–673
15. Simon G, Ormel J, VonKorff M, et al. Health care costs associated with depressive and anxiety disorders in primary care. Am J Psychiatry 1995;152:352–357
16. Miller IW, Keitner GI, Schatzberg AF, et al. The treatment of chronic depression, pt 3: psychosocial functioning before and after treatment with sertraline or imipramine. J Clin Psychiatry 1998;59:608–619
17. Mintz J, Mintz LI, Arruda MJ, et al. Treatments of depression and the functional capacity to work. Arch Gen Psychiatry 1992;49:761–768
18. Broadhead WE, Blazer DG, George LK, et al. Depression, disability days, and days lost from work in a prospective epidemiologic survey. JAMA 1990;264:2524–2528
19. Von Korff M, Ormel J, Katon W, et al. Disability and depression among high utilizers of health care: a longitudinal analysis. Arch Gen Psychiatry 1992;49:91–100
20. Kaplan R. Using general measures of health related quality of life to assess mental health outcomes. In: Miller N, Magruder K, eds. Cost-Effectiveness of Psychotherapy: A Guide for Practitioners, Researchers and Policy-Makers. New York, NY: John Wiley & Sons; 1999
21. Anton S, Revicki D. The use of decision analysis in the pharmacoeconomic evaluation of an antidepressant: a cost-effectiveness study of nefazodone. Psychopharmacol Bull 1995;31:249–258
22. Fryback DG, Dasbach EJ, Klein R, et al. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. Med Decis Making 1993;13:89–102
23. Ballenger J. Clinical guidelines for establishing remission in patients with depression and anxiety. J Clin Psychiatry 1999;60(suppl 22):29–34
24. Nierenberg AA, Wright EC. Evolution of remission as the new standard in the treatment of depression. J Clin Psychiatry 1999;60(suppl 22):7–11
25. Frank E, Prien RF, Jarrett RB, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. Arch Gen Psychiatry 1991;48:851–855
26. Keller MB, Lavori PW, Friedman B, et al. The longitudinal interval follow-up evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. Arch Gen Psychiatry 1987;44: 540–548
27. Quitkin F, Rabkin JD, Markowitz JC, et al. Use of pattern analysis to identify true drug response: a replication. Arch Gen Psychiatry 1987;44:259–264
28. Thase ME. Introduction: defining remission in patients treated with antidepressants. J Clin Psychiatry 1999;60(suppl 22):3–6
29. Lave J, Frank RG, Schulberg H, et al. Cost-effectiveness of treatments for depression in primary care practice. Arch Gen Psychiatry 1998;55:645–651
30. Thase ME, Entsuah AR, Rudolph RL. Remission rates during treatment with venlafaxine or selective serotonin reuptake inhibitors. Br J Psychiatry 2001;178:234–241
31. Silverstone PH, Ravindran A, for the Venlafaxine XR 360 Study Group. Once-daily venlafaxine extended release (XR) compared with fluoxetine in outpatients with depression and anxiety. J Clin Psychiatry 1999;60: 22–28
32. Clerc GE, Ruimy P, Verdeau-Palles J. A double-blind comparison of venlafaxine and fluoxetine in patients hospitalized for major depression and melancholia. The Venlafaxine French Inpatient Study Group. Int Clin Psychopharmacol 1994;9:139–143
33. Dierick M, Ravizza L, Realini R, et al. A double-blind comparison of venlafaxine and fluoxetine for treatment of major depression in outpatients. Prog Neuropsychopharmacol Biol Psychiatry 1996;20:57–71
34. 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
35. Rudolph RL, Entsuah R, Aguiar L. Early onset of antidepressant activity of venlafaxine compared with placebo and fluoxetine in outpatients in a double-blind study [abstract]. Eur Neuropsychopharmacol 1998;8 (suppl 2):S142
36. Salinas E, for the Venlafaxine XR 367 Study Group. Once-daily venlafaxine extended release XR venlafaxine versus paroxetine in outpatients

with major depression [abstract]. Biol Psychiatry 1997;42(suppl 1):244S
37. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised. Washington, DC: American Psychiatric Association; 1987
38. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington, DC: American Psychiatric Association; 1994
39. Stewart JW, Quitkin FM, McGrath PJ, et al. Use of pattern analysis to predict differential relapse of remitted patients with major depression during one year of treatment with fluoxetine or placebo. Arch Gen Psychiatry 1998;55:334–343
40. Guyatt G, Jaeschke R. Measurements in clinical trials: choosing the appropriate approach. In: Spilker B, ed. Quality of Life Assessments in Clinical Trials. New York, NY: Raven Press; 1990:37–46
41. Jaeschke R, Guyatte G. How to develop and validate a new quality of life instrument. In: Spilker B, ed. Quality of Life Assessments in Clinical Trials. New York, NY: Raven Press; 1990:47–57
42. Lydick E, Epstein RS. Interpretation of quality of life changes. Qual Life Res 1993;2:221–226
43. Juniper EF, Guyatt GH, Willan A, et al. Determining a minimal important change in a disease-specific Quality of Life Questionnaire. J Clin Epidemiol 1994;47:81–87
44. Testa MA. Interpreting quality-of-life clinical trial data for use in the clinical practice of antihypertensive therapy. J Hyperten Suppl 1987;5:S9–S13
45. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27(3, suppl):S178–S189
46. Fletcher A, Gore S, Jones D, et al. Quality of life measures in health care, 2: design, analysis, and interpretation. BMJ 1992;305:1145–1148
47. Lenderking WR, Tennen H, Nackley JF, et al. The effects of venlafaxine on social activity level in depressed outpatients. J Clin Psychiatry 1999;60:157–163
48. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, NY: Academic Press; 1977
49. Rickels K, Derivan A, Entsuah R, et al. Rapid onset of antidepressant activity with venlafaxine treatment. Depression 1995;3:146–153
50. Sculpher MJ, Buxton MJ. The episode-free day as a composite measure of effectiveness: an illustrative economic evaluation of formoterol versus salbutamol in asthma therapy. Pharmacoeconomics 1993;4:345–352
51. Freeman EW, DeRubeis RJ, Rickels K. Reliability and validity of a daily diary for premenstrual syndrome. Psychiatry Res 1996;65:97–106
52. Ninan PT, McElroy SL, Kane CP, et al. Placebo-controlled study of fluvoxamine in the treatment of patients with compulsive buying. J Clin Psychopharmacol 2000;20:362–366
53. Diggle P, Liang K, Zeger S. Analysis of Longitudinal Data. Oxford, England: Clarendon Press; 1994

Appendix 1. Depression-Free Days: The Construct

Depression-free days are derived from weekly 17-item Hamilton Rating Scale for Depression (HAM-D-17) scores. The derivation involves the following steps:

(i) Calculate average intervisit HAM-D score ($HAM\text{-}D_{i,i+1}$):

$$\text{Average intervisit HAM-D}_{i,i+1} = \frac{\text{HAM-D}_i + \text{HAM-D}_{i+1}}{2}$$

where i = visit number = 0 (baseline), 1, 2, …8 (end of analysis period)

(ii) Adjust average $HAM\text{-}D_{i,i+1}$ to minimum-maximum range of 7–15:

If average $HAM\text{-}D_{i,i+1} \leq 7$, then $HAM\text{-}D_{i,i+1} = 7$
If average $HAM\text{-}D_{i,i+1} \geq 15$, then $HAM\text{-}D_{i,i+1} = 15$

(iii) Calculate relative deviation of average $HAM\text{-}D_{i,i+1}$ from maximum HAM-D:

$$\text{Relative deviation of average HAM-D from maximum} = \frac{15 - \text{average (HAM-D}_{i,i+1})}{15 - 7}$$

This should always be between 0 and 1, given step ii.

(iv) Multiply by number of actual days between assessments:

$$\text{Depression-free days (DFD}_{i,i+1}) = \frac{15 - \text{average (HAM-D}_{i,i+1})}{15 - 7} \times 7^*$$

*(or actual number of days between visits i and i + 1)
This should then yield a possible range of 0 to (7 or actual number of days between visits).

(v) Sum (iv) over all intervisit average HAM-D scores to get depression-free days over entire study:

$$\text{Depression-free days over analysis period} = \sum_{i=0}^{i=7} \text{DFDs}_{i,i+1}$$

(possible range of DFDs over the 8 weeks of the analysis = 0 – 56 days).

Note: In the secondary DFDs construct, the above derivation is modified by substituting 22 for 15 in all steps. In the tertiary DFDs construct (representing elimination of residual symptoms), the above derivation is modified by successively substituting the values 6, 5, 4, 3, 2, and 1 for 7 in the above formula; results for the lower threshold of 1 are presented in Table 3.

