

# Effect of *Hypericum perforatum* (St John's Wort) in Major Depressive Disorder

## A Randomized Controlled Trial

### Hypericum Depression Trial Study Group

**H**YPERICUM PERFORATUM (St John's wort) is widely used to treat depression, sometimes in an attempt to avoid adverse effects associated with prescription antidepressants. One meta-analysis in 1996 concluded that hypericum is superior to placebo for treatment of mild to moderate depression.<sup>1</sup> Subsequent studies have found hypericum to be comparable to active controls, such as amitriptyline,<sup>2</sup> imipramine,<sup>3-5</sup> and fluoxetine,<sup>6</sup> and superior to placebo.<sup>4,7</sup> Some studies suggest that it may be an effective treatment for moderately severe depression.<sup>3,4</sup> Others have been unable to differentiate hypericum from placebo.<sup>8,9</sup>

Important issues have been raised regarding existing studies, including limited information about use in clinically defined major depression, lack of placebo-controlled trials that have included a selective serotonin reuptake inhibitor arm, and absence of controlled data for continuation treatment. Concern has been raised about adverse interactions of hypericum with certain drugs.<sup>10,11</sup> Most hypericum in the United States is consumed without physician consultation. Even though many patients prefer to avoid the use of medications with adverse effects, there is a risk that people with clinically significant depression may self-medicate with

**Context** Extracts of *Hypericum perforatum* (St John's wort) are widely used for the treatment of depression of varying severity. Their efficacy in major depressive disorder, however, has not been conclusively demonstrated.

**Objective** To test the efficacy and safety of a well-characterized *H perforatum* extract (LI-160) in major depressive disorder.

**Design and Setting** Double-blind, randomized, placebo-controlled trial conducted in 12 academic and community psychiatric research clinics in the United States.

**Participants** Adult outpatients (n=340) recruited between December 1998 and June 2000 with major depression and a baseline total score on the Hamilton Depression Scale (HAM-D) of at least 20.

**Interventions** Patients were randomly assigned to receive *H perforatum*, placebo, or sertraline (as an active comparator) for 8 weeks. Based on clinical response, the daily dose of *H perforatum* could range from 900 to 1500 mg and that of sertraline from 50 to 100 mg. Responders at week 8 could continue blinded treatment for another 18 weeks.

**Main Outcome Measures** Change in the HAM-D total score from baseline to 8 weeks; rates of full response, determined by the HAM-D and Clinical Global Impressions (CGI) scores.

**Results** On the 2 primary outcome measures, neither sertraline nor *H perforatum* was significantly different from placebo. The random regression parameter estimate for mean (SE) change in HAM-D total score from baseline to week 8 (with a greater decline indicating more improvement) was -9.20 (0.67) (95% confidence interval [CI], -10.51 to -7.89) for placebo vs -8.68 (0.68) (95% CI, -10.01 to -7.35) for *H perforatum* ( $P=.59$ ) and -10.53 (0.72) (95% CI, -11.94 to -9.12) for sertraline ( $P=.18$ ). Full response occurred in 31.9% of the placebo-treated patients vs 23.9% of the *H perforatum*-treated patients ( $P=.21$ ) and 24.8% of sertraline-treated patients ( $P=.26$ ). Sertraline was better than placebo on the CGI improvement scale ( $P=.02$ ), which was a secondary measure in this study. Adverse-effect profiles for *H perforatum* and sertraline differed relative to placebo.

**Conclusion** This study fails to support the efficacy of *H perforatum* in moderately severe major depression. The result may be due to low assay sensitivity of the trial, but the complete absence of trends suggestive of efficacy for *H perforatum* is noteworthy.

JAMA. 2002;287:1807-1814

www.jama.com

hypericum rather than receive effective medication or psychotherapy.

This placebo-controlled study was designed to expand on previous trials by studying outpatients with well-defined

**Members of the Hypericum Depression Trial Study Group and Financial Disclosures** are listed at the end of this article.

**Corresponding Author and Reprints:** Jonathan R. T. Davidson, MD, Duke University Medical Center, Box 3812, Durham, NC 27710 (e-mail: jonathan.davidson@duke.edu).

See also pp 1840 and 1853.

major depression of moderate severity and included a 4-month continuation phase and sertraline as an active comparator to calibrate the trial's validity. The main hypothesis tested whether hypericum would be superior to placebo after 8 weeks of treatment.

## METHODS

The study was a randomized, double-blind, parallel-group, 8-week, outpatient trial of hypericum, sertraline, or placebo treatment for major depressive disorder, followed by up to 18 weeks of double-blind continuation treatment in participants meeting response criteria at 8 weeks.

### Patients

Outpatients meeting *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)* criteria for major depressive disorder<sup>12</sup> were recruited from 12 academic or community clinics between December 1998 and June 2000. Major depressive disorder was diagnosed with the modified Structured Clinical Interview for Axis I DSM-IV disorders (SCID-Hypericum).<sup>13</sup>

Inclusion criteria were age at least 18 years; current diagnosis of major depression; minimum total score of 20 on the 17-item Hamilton Depression (HAM-D)<sup>14</sup> scale and a maximum score of 60 on the Global Assessment of Functioning (GAF)<sup>12</sup> at screening and baseline following a 1-week, single-blind, placebo run-in; no more than a 25% decrease in HAM-D total score between screening and baseline; capacity to give informed consent and follow study procedures; and identification of a close personal contact to be notified if warranted by clinical concerns.

Exclusion criteria were a score above 2 on the HAM-D suicide item; attempted suicide in the past year or current suicide or homicide risk; being pregnant, planning pregnancy, breastfeeding, or not using medically acceptable birth control; clinically significant liver disease or liver enzyme levels elevated to at least twice the upper normal limit; serious unstable medical illness; history of seizure disorder; SCID diagnoses indi-

cating alcohol or other substance-abuse disorder within the past 6 months or lifetime diagnoses of schizophrenia, schizoaffective or other psychotic disorder, bipolar disorder, panic disorder, or obsessive-compulsive disorder; history of psychotic features of affective disorder; evidence of untreated or unstable thyroid disorder; no response to at least 2 adequate trials of antidepressants in any depressive episode; daily use of hypericum or sertraline for at least 4 weeks within the past 6 months; current use of other psychotropic drugs, other medicines, dietary supplements, natural remedies, or botanical preparations with psychoactive properties; use of investigational drugs within 30 days of baseline or of other psychotropic drugs within 21 days of baseline (within 6 weeks for fluoxetine); allergy or hypersensitivity to study medications; positive urine drug screen; introduction of psychotherapy within 2 months of enrollment or any ongoing psychotherapy specifically designed to treat depression; and mental retardation or cognitive impairment.

### Study Design

Patients provided written informed consent, and the institutional review board approved the protocol at each site. Patients who remained eligible after a 1-week placebo run-in were randomly assigned to receive 1 of 3 treatments in a 1:1:1 ratio within permuted blocks of size 3 and 6 within site by sex strata. Sites telephoned a 24-hour randomization service for computer-generated treatment assignment. Drug kits were designed to be indistinguishable for the 3 treatments at each dose level.

Patients were assessed weekly or bi-weekly until week 8. Patients who fully or partially responded during these 8 weeks (ie, the acute phase) could enter the continuation phase, with visits at weeks 10, 14, 18, 22, and 26. The HAM-D, GAF, Clinical Global Impressions Scales<sup>15</sup> for severity (CGI-S) and improvement (CGI-I) and Beck Depression Inventory (BDI)<sup>16</sup> were assessed at all visits. The Sheehan Disability Scale (SDS)<sup>17</sup> was completed at baseline and weeks 8 and 26. As a way

to evaluate blinding at weeks 8 and 26, clinicians and patients indicated their beliefs about treatment assignment.

Drug accountability, concomitant therapies, vital signs, and self- and physician-rated symptom reports were assessed at every visit. Blood chemistry and hematologic tests and electrocardiography and physical examinations were performed at screening and at weeks 8 and 26. Urine toxicology was performed at screening.

Columbia University biometric staff trained all raters in the use of the SCID and HAM-D. For reliable scoring of the CGI, raters scored case vignettes before the study. Throughout, audiotapes of SCID and HAM-D interviews and of the medication-management sessions were audited by the coordinating center for quality and adherence to the guidelines specified in the operations manual.

### Study Medications

Hypericum and matching placebo were provided by Lichtwer Pharma (Berlin, Germany); sertraline and matching placebo were provided by Pfizer Inc (New York, NY). The Lichtwer extract (LI-160) was selected for its well-characterized features and literature supporting its possible efficacy in depression.<sup>1,18,19</sup> The extract was standardized to between 0.12% and 0.28% hypericin, and the entire supply came from one batch. The study was conducted under an investigational new drug application filed by the manufacturer.

Medications were given 3 times daily. During the run-in period, patients received placebo tablets in a single-blind fashion. At baseline, patients were randomly assigned to receive hypericum (900 mg/d), sertraline (50 mg/d), or placebo. Daily hypericum, sertraline, or placebo doses could be increased to 1200 mg, 75 mg, or placebo equivalent, respectively, after weeks 3 or 4 and to 1500 mg, 100 mg, or placebo equivalent at week 6 if the CGI-S score was 4 (moderately ill) or more at week 3, or 3 (mildly ill) or more at weeks 4 or 6. After week 8, those eligible to continue could re-

ceive maximum daily doses of 1800 mg, 150 mg, or placebo equivalent, respectively. Medication was dispensed in blister packets in double-dummy fashion. Doses could be held or reduced for adverse effects. For insomnia, zolpidem (5 to 10 mg) was permitted up to twice weekly during weeks 1 and 2 and up to 6 times total during continuation.

### Efficacy End Points

The prospectively defined primary efficacy measures were the change in the 17-item HAM-D total score from baseline to week 8 and the incidence of full response at week 8 or early study termination. Full response was defined as a CGI-I score of 1 (very much improved) or 2 (much improved) and a HAM-D total score of 8 or less. Partial response was defined as a CGI-I score of 1 or 2, a decrease in the HAM-D total score from baseline of at least 50%, and a HAM-D total score of 9 to 12. Secondary end points comprised the GAF, CGI, BDI, and SDS scores.

After week 8, relapse was defined as a HAM-D score of 20 or more and a CGI-S score of 4 or more at 2 consecutive visits. Serious suicidal ideation or the development of psychosis also served as grounds for removal from the study and prompt clinical assessment.

### Safety Assessments

Any symptom or sign that appeared or became worse after baseline was considered an adverse event. Adverse events were elicited and recorded by the study physician at each visit, based on patient interview and on a 44-item checklist completed by the patient and expanded from an earlier scale.<sup>20</sup>

### Compliance

Patients were deemed noncompliant if they had taken less than 80% of the prescribed medication, according to pill counts at each follow-up visit.

### Statistical Analysis

The principal comparison was between the hypericum and placebo groups. Sertraline served as an active comparator to evaluate the study's sen-

sitivity. Sample-size calculations were based on detecting a difference in full-response rates at 8 weeks, assuming full-response rates of 55% for hypericum and 35% for placebo. Accordingly, a sample size of 336 patients (112 per group) was specified to ensure 85% power with a type I error rate of 5% (2-sided). The sample-size calculation assumed no interactions of treatment with site or sex, the blocking factors for randomization.

The primary analysis was according to assignment at randomization. However, a systematic review of all protocol deviations in patient enrollment, as indicated by the database, was undertaken before the study was unblinded, and patients who did not meet the HAM-D total score of at least 20 entry criteria ( $n=2$ ) were excluded from the efficacy analyses as recommended by the scientific advisors. These ineligibilities resulted from mistakes in summing the 17-item scores.

Treatment differences in full-response rates were assessed with Wald  $\chi^2$  statistics from logistic regression, with fixed effects for treatment, site, sex, and baseline HAM-D total score. Treatment differences in the change in HAM-D total score from baseline to week 8 were evaluated through a random-coefficient regression model. The longitudinal scores at baseline and weeks 1 through 8 (all available acute-phase data) were modeled as a linear function of fixed effects for treatment, site, sex, study week (linear), and treatment by study week, with random intercept and slope over time for each patient. Under the assumptions of this model, tests of treatment differences for the change in HAM-D total score from baseline to week 8 are equivalent to tests of treatment differences in the linear trends or slopes with time.

These analyses on full response and change in the HAM-D total score were specified in the final protocol as primary in assessment of acute-phase efficacy. Secondary analyses included random-coefficient regression models on secondary outcomes as described above for HAM-D, similar modeling on pri-

mary and secondary outcomes but restricted to patients completing the acute phase (completer analysis), and analysis of covariance models on primary and secondary outcomes using the last available acute-phase measurement (last observation carried forward). The analysis-of-covariance models included effects for treatment, site, sex, and the respective baseline scores.

Among the 3 treatment comparisons, those of hypericum vs placebo and sertraline vs placebo were of interest a priori, and their *P* values (2-sided) are presented. In addition, for efficacy measures, the nominal significance level for the hypericum vs sertraline contrast is noted in the text if  $P < .05$  when the 3 treatment groups differed overall (2 *df*) using a type I error of .05.

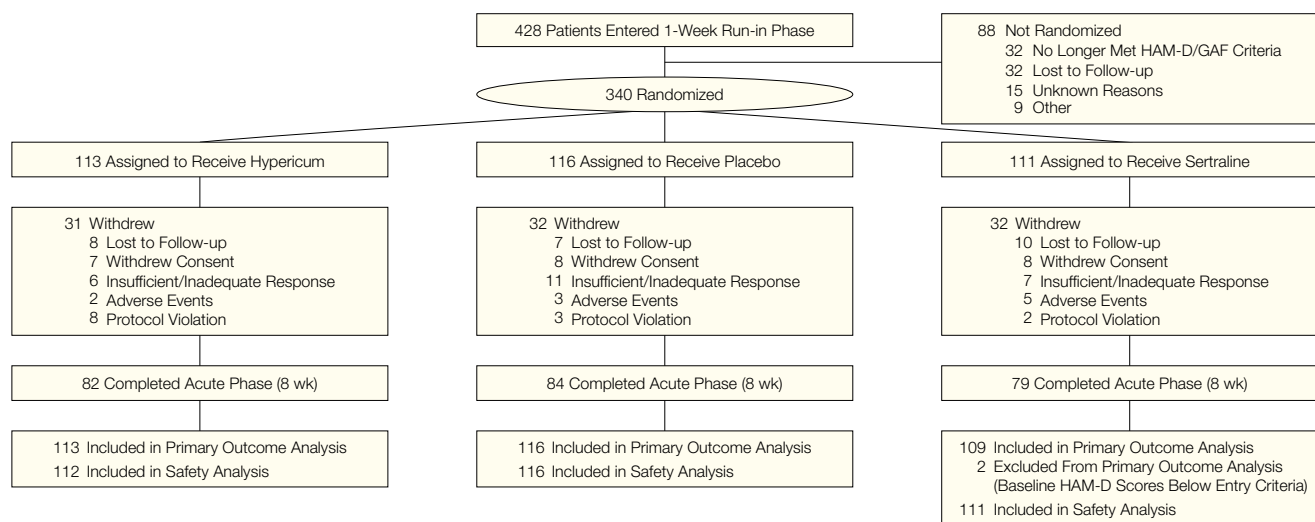
Simple tests for treatment differences included  $\chi^2$  and Fisher exact tests for categorical variables, Kruskal-Wallis/Wilcoxon-Mann-Whitney tests for ordinal and continuous measurements, and log-rank tests for time to events. These tests were applied to baseline characteristics, protocol deviations, adverse events, attrition, compliance, treatment beliefs, and maintenance of response during continuation. In addition, the nonparametric methods were used on the efficacy measures to substantiate results that rely on distributional assumptions. The consistency of the data with the parametric assumptions was checked for the primary analyses.

All analyses were performed with SAS version 6.12 software (SAS Institute Inc, Cary, NC). The PROC MIXED procedure was used for the longitudinal data models.

## RESULTS

In all, 428 patients entered the run-in phase, and 340 were randomized (FIGURE 1). No differences were noted between treatment groups at baseline (TABLE 1). In addition, with regard to severity of depression, the numbers of patients judged to be mild, moderate, marked, and severe were 3, 261, 70, and 6, respectively, according to the CGIs. Baseline total HAM-D scores ranged from 18 to 33.

**Figure 1.** Disposition of the Study Patients During the Acute Phase



HAM-D indicates Hamilton Depression scale; GAF, Global Assessment of Functioning.

**Table 1.** Baseline Demographic and Clinical Characteristics\*

Characteristic	Hypericum (n = 113)	Placebo (n = 116)	Sertraline (n = 111)
Age, mean (SD), y	43.1 (13.5)	40.1 (12.2)	43.9 (13.9)
Sex, No. (%)			
Male	40.0 (35.4)	39.0 (33.6)	37.0 (33.3)
Female	73.0 (64.6)	77.0 (66.4)	74.0 (66.7)
Race, No. (%)			
Black	10 (8.8)	12 (10.3)	13 (11.7)
White	87 (77.0)	88 (75.9)	82 (73.9)
Hispanic	10 (8.8)	9 (7.8)	8 (7.2)
Other	6 (5.3)	7 (6.0)	8 (7.2)
Duration of current depression, No. (%)			
<6 mo	32.0 (28.6)	41.0 (35.3)	43.0 (38.7)
6-24 mo	42.0 (37.5)	42.0 (36.2)	31.0 (27.9)
>24 mo	38.0 (33.9)	33.0 (28.4)	37.0 (33.3)
HAM-D total score, mean (SD)	23.1 (2.7)	22.7 (2.7)	22.5 (2.5)
GAF score, mean (SD)	53.2 (5.0)	54.1 (4.3)	53.6 (4.7)
BDI total score, mean (SD)	21.6 (8.5)	21.2 (8.8)	21.2 (8.2)
SDS total score, mean (SD)	15.8 (8.3)	15.3 (8.1)	15.4 (7.9)
CGI severity score, mean (SD)	4.3 (0.5)	4.2 (0.5)	4.2 (0.5)

\*HAM-D indicates Hamilton Depression scale; GAF, Global Assessment of Functioning; BDI, Beck Depression Inventory; SDS, Sheehan Disability Scale; and CGI, Clinical Global Impressions.

There were similar proportions of patients among the treatment groups discontinued before week 8: 27% for hypericum (n = 31), 28% for placebo (n = 32), and 29% for sertraline (n = 32). Likewise, time to early discontinuation did not differ significantly (P = .91, log-rank test), although 17 of the 32 dropouts in the sertraline group (53%) occurred during the first 2 weeks vs 5

of 32 dropouts in the placebo group (16%) and 11 of 31 dropouts in the hypericum group (35%).

Of the 340 acute-phase subjects, 245 (72%) completed 8 weeks, 129 entered the continuation phase, and 79 completed continuation. There were no treatment differences in attrition during continuation. Nine of the 129 patients entering continuation did not

meet response criteria. These patients and the 2 who were ineligible for the acute phase were excluded from efficacy analysis in the continuation phase.

The mean (SD) highest daily dose prescribed was 1299 (243) mg (95% confidence interval [CI], 1254-1344 mg) for hypericum and 75 (21) mg (95% CI, 71-79 mg) for sertraline during the acute phase and 1382 (284) mg (95% CI, 1292-1473 mg) for hypericum and 89 (32) mg (95% CI, 80-98 mg) for sertraline during continuation. Fewer patients in the sertraline group achieved the highest daily dose level during the acute phase (36% compared with 54% for hypericum and 54% for placebo; P = .005, Kruskal-Wallis test). Similar proportions of each treatment group required dose reductions during the acute phase because of adverse events (hypericum, 4%; placebo, 5%; and sertraline, 7%).

**Efficacy Outcomes**

The HAM-D total scores throughout the 8-week trial are summarized by treatment group in FIGURE 2. The random-coefficient regression analysis on the longitudinal HAM-D total scores detected a downward linear trend with time (F<sub>1,263</sub> = 565.2; P < .001) and gen-

eral differences in scores by site ( $F_{12,328}=4.56$ ;  $P<.001$ ) and sex (lower for men;  $F_{1,326}=8.97$ ;  $P=.003$ ). Linear trends with time did not differ significantly by treatment (hypericum vs placebo:  $F_{1,265}=0.30$  and  $P=.59$ ; sertraline vs placebo:  $F_{1,264}=1.83$  and  $P=.18$ ), and no interactions were found between treatment and site or sex. Model estimates for the mean (SE) change in HAM-D total score (week 8 minus baseline) were  $-8.68$  (0.68) for hypericum,  $-9.20$  (0.67) for placebo, and  $-10.53$  (0.72) for sertraline. According to the model estimates for the difference in slopes between sertraline and placebo and the variance estimate for the random slope coefficient, the sertraline effect size was 0.24.

Full response rates at acute-phase exit did not differ between placebo and either hypericum ( $P=.21$ ) or sertraline ( $P=.26$ ) (TABLE 2). Patients with a lower HAM-D total score at baseline had a higher rate of full response ( $P=.002$ ), which did not differ by treatment group. No differences were noted by site or sex. No interactions were found for treatment by site, sex, or baseline HAM-D total score. The ordinal distribution of responses (full, partial, or none) differed significantly across treatments ( $P=.04$ , Kruskal-Wallis test), with more partial responders among the sertraline patients and more full responders among the placebo patients. Findings were similar for patients completing the acute phase, with percentages (full, partial, or no response) of 30.5%, 17.1%, and 52.4% for hypericum, 41.6%, 14.3%, and 44.0% for placebo, and 31.2%, 29.9%, and 39.0% for sertraline, respectively.

Other acute-phase outcomes did not differ between hypericum and placebo (TABLE 3). Sertraline was superior to placebo on the CGI-I score at week 8 ( $F_{1,254}=6.36$ ;  $P=.02$ ) and showed a general trend toward better outcomes. In post hoc analysis, sertraline proved superior to hypericum on the CGI-I scale ( $F_{1,252}=7.91$ ;  $P=.01$ ).

The numbers of patients who entered continued therapy for hypericum, placebo, and sertraline were 38, 42,

and 49, respectively; for those completing treatment, 24, 27, and 28, respectively. Efficacy assessment during continuation excluded the 2 patients not meeting the HAM-D entry criteria at baseline and 9 patients who began continuation without meeting the full or partial response criteria. The HAM-D total score means (SDs) were 6.7 (3.5), 6.2 (3.0), and 6.9 (3.6) at entry to continuation and 6.6 (4.5), 5.3 (5.2), and 6.7 (4.9) at week 26, respectively, by treatment. Only 1 patient (receiving hypericum) relapsed during continuation.

**Safety Evaluation**

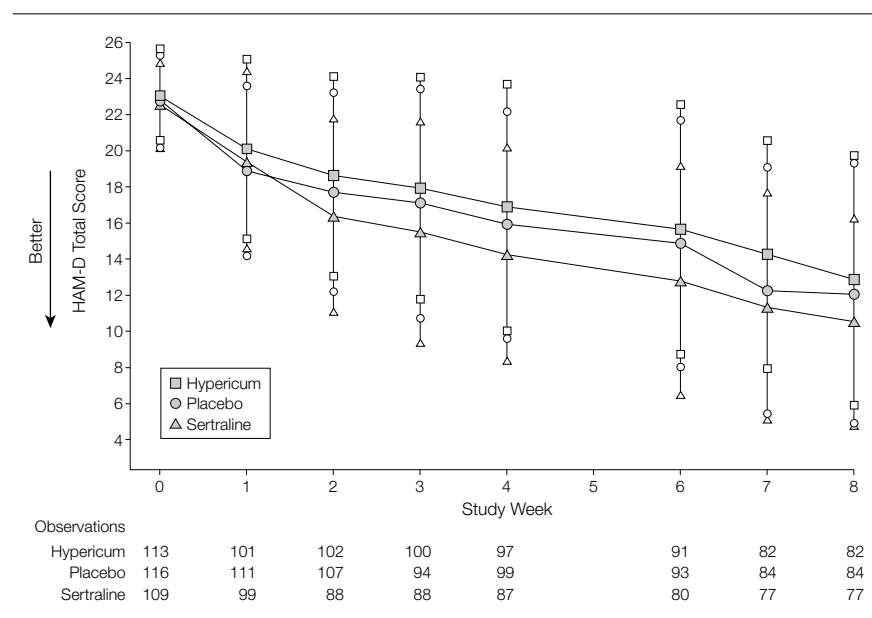
Hypericum and sertraline were associated with more acute-phase adverse

events than placebo. TABLE 4 displays those that differed significantly. Analyses of data from the continuation phase yielded similar results. Rates of diarrhea, nausea, and sweating (sertraline); anorgasmia (sertraline and hypericum); and frequent urination and swelling (hypericum) all were higher than those of placebo. Forgetfulness was less common with sertraline than with placebo. No serious adverse events occurred.

**Assessment of Blindness to Treatment**

At the end of 8 weeks, the proportion of patients guessing their treatment correctly was 55% for sertraline, 29% for

**Figure 2.** The 17-Item Hamilton Depression (HAM-D) Scale Total Score by Acute Phase Week



Graphed values are means with vertical bars extending to one SD.

**Table 2.** Clinical Response Rates\*

Response, No. (%)	Hypericum (n = 113)	Placebo (n = 116)	Sertraline (n = 109)†
Any response	43.0 (38.1)	50.0 (43.1)	53.0 (48.6)
Full response‡	27.0 (23.9)	37.0 (31.9)	27.0 (24.8)
Partial response§	16.0 (14.2)	13.0 (11.2)	26.0 (23.9)
No response	70.0 (61.9)	66.0 (56.9)	56.0 (51.4)

\*No significant differences were noted between hypericum and placebo ( $P=.21$ ) or between sertraline and placebo ( $P=.26$ ) in the log odds of full response with adjustment for site, sex, and baseline Hamilton Depression (HAM-D) scale total score through logistic regression. Overall differences in the response distribution ( $P=.04$ , Kruskal-Wallis test) were noted because of the higher number of partial responders with sertraline and full responders with placebo. †Excludes 2 patients who had baseline HAM-D scores below the cutoff required for enrollment.

‡Indicates a HAM-D score of 8 or less and a Clinical Global Impressions improvement (CGI-I) score of 1 or 2. §Indicates a HAM-D score of 9 to 12, inclusive, a 50% reduction in HAM-D score, and a CGI-I score of 1 or 2.

**Table 3.** Summary Measures of Acute-Phase Efficacy\*

Summary Measure‡	Model Estimates, Mean (SE) [95% Confidence Interval]†			P Value†	
	Hypericum (n = 113)	Placebo (n = 116)	Sertraline (n = 109)§	Hypericum vs Placebo	Sertraline vs Placebo
HAM-D total score (week 8 – baseline)	-8.68 (0.68) [-10.01 to -7.35]	-9.20 (0.67) [-10.51 to -7.89]	-10.53 (0.72) [-11.94 to -9.12]	.59	.18
GAF score (week 8 – baseline)	12.31 (1.22) [9.92 to 14.70]	11.76 (1.19) [9.43 to 14.09]	14.16 (1.27) [11.67 to 16.65]	.75	.17
BDI total score (week 8 – baseline)	-7.84 (0.91) [-9.62 to -6.06]	-6.83 (0.90) [-8.59 to -5.07]	-8.75 (0.95) [-10.61 to -6.89]	.43	.15
SDS total score (acute exit – baseline)	-3.34 (0.69) [-4.69 to -1.99]	-3.19 (0.68) [-4.52 to -1.86]	-4.01 (0.70) [-5.38 to -2.64]	.88	.39
CGI-S score (week 8 – baseline)	-1.07 (0.12) [-1.31 to -0.83]	-1.19 (0.12) [-1.43 to -0.95]	-1.47 (0.12) [-1.71 to -1.23]	.48	.11
CGI-I score (week 8)	2.50 (0.11) [2.28 to 2.72]	2.45 (0.11) [2.23 to 2.67]	2.04 (0.12) [1.80 to 2.28]	.75	.02

\*See Table 1 for definitions of the scoring tools. CGI-S indicates Clinical Global Impressions severity score; CGI-I, Clinical Global Impressions improvement score.  
 †Values are linear model estimates and P values adjusted for site and sex. HAM-D total score, GAF score, BDI total score, and CGI-S score statistics were based on modeling the longitudinal measures at week 0 and weeks 1 through 8 as a linear function of site, sex, treatment, week, and treatment by week, with a random intercept and slope over time for each subject. A similar model was used for CGI-I, but there was no baseline measurement in the longitudinal stream. The SDS was measured at baseline and acute-phase exit. Change was modeled as a linear function of the baseline measure, site, sex, and treatment. The means for CGI-I and SDS are least square.  
 ‡Lower scores denote greater improvement, except for the GAF.  
 §Two sertraline patients had HAM-D scores below the cutoff of 20 required for enrollment. These patients were excluded from efficacy analysis.

**Table 4.** Adverse Events That Differed Significantly by Treatment During the Acute Phase\*

Event, No. (%)	Hypericum (n = 112)‡	Placebo (n = 116)	Sertraline (n = 111)	P Value†	
				Hypericum vs Placebo	Sertraline vs Placebo
Diarrhea	23 (21)	22 (19)	42 (38)	.81	.003
Nausea	21 (19)	24 (21)	41 (37)	.78	.02
Anorgasmia	28 (25)	16 (14)	35 (32)	.04	.002
Forgetfulness	28 (25)	26 (22)	13 (12)	.75	.04
Frequent urination	30 (27)	13 (11)	23 (21)	.003	.06
Sweating	20 (18)	14 (12)	32 (29)	.21	.003
Swelling	21 (19)	9 (8)	9 (8)	.02	.90

\*No attributions made with respect to the relationship to treatment.  
 †Wilcoxon-Mann-Whitney test for difference in distribution of severity between treatments (absent, mild, moderate, or severe).  
 ‡Excluding 1 patient who lacked a symptom report.

hypericum, and 31% for placebo ( $P = .02$  for differences between treatment groups). Correct guesses for clinicians totaled 66% for sertraline, 29% for hypericum, and 36% for placebo ( $P = .001$  for differences between treatment groups). The change (mean) in HAM-D total score from baseline to week 8 did not differ for patients who were in the sertraline group and either had guessed the correct treatment (-11.6; 95% CI, -13.1 to -10.1) or had not (-11.9; 95% CI, -13.9 to -9.9).

**COMMENT**

As with 2 other trials,<sup>8,9</sup> we have found no evidence for a superior effect of hypericum relative to placebo. Neither hypericum nor sertraline could be differ-

entiated from placebo on the primary efficacy measures. Although the efficacy of sertraline was demonstrated on the secondary CGI-I measure, resulting on average in much improvement, hypericum had no efficacy on any measure. Although not designed to compare sertraline with hypericum, the study showed superiority of sertraline on the CGI-I. Responders who entered continuation treatment maintained their improvement equally in each treatment group.

The overall effect size for sertraline on the HAM-D total score was 0.24, which is consistent with reported effect sizes for standard antidepressants,<sup>21,22</sup> while on the CGI-I, it was 0.41. These findings can also be observed in the con-

text of 3 other sertraline studies<sup>23-25</sup> that yielded effect sizes of 0.31, 0.33, and 0.45 for the drug relative to placebo on the HAM-D change from baseline to last observation in all study patients.

Adverse effects for sertraline were consistent with its profile, while for hypericum, more frequent anorgasmia, swelling, and urination were noted relative to placebo, although these were mild and the multiple comparisons may have produced spurious associations.

When a new treatment cannot be distinguished from placebo, it is important to determine whether a drug of known efficacy would have been proven effective in that sample. Failure of established antidepressants to show such superiority occurs in up to 35% of trials,<sup>26,27</sup> which illustrates the difficulties plaguing randomized placebo-controlled trials in this population. We should, therefore, consider some of the factors that might have contributed to our results. One concern is a high placebo response rate, but this was not unusually high in our sample and is therefore an unlikely explanation.

Addressing specific issues relevant to sertraline, we note the following. Although a dose-effect relationship within the therapeutic range (50-200 mg/d) has not been demonstrated,<sup>28-30</sup> one may wonder whether the study dose limitation (up to 100 mg/d in the 8-week trial)

was too restrictive and whether even the highest prescribed dose was administered for an inadequate duration. The protocol dose regimen was chosen on the basis of extensive discussion by all parties concerned in the study design and oversight as the best compromise for ensuring effective treatment while minimizing the incidence of dose-related adverse events. In fact, only 36% of sertraline patients had their dose maximized to 100 mg/d compared with 54% for hypericum or placebo. There were more partial responders to sertraline (23.9%) than to hypericum (14.2%) or placebo (11.2%;  $P = .03$ ). Analyses of the sertraline patients eligible for a dose increase revealed no association between lack of dose increase and presence of adverse events. Thus, it appears that, in this trial, clinicians tended not to increase the sertraline dose for patients with partial response, electing instead to allow more time with the same dose. On the matter of dosing, if any protocol bias existed at all, it would favor hypericum, which could be dosed to the maximum of its permissible range, whereas the maximum permitted dose of sertraline was only 50% of its highest recommended amount.

The study was adequately powered to detect moderate effect sizes (ie, at least 0.40). The observed sertraline effect size was small (0.24) on the HAM-D total score and moderate (0.41) on the CGI-I; hence, the lack of statistical significance on the primary outcome measure.

For hypericum, 2 issues are relevant. Although the hyperforin content of this batch was 3.1%, the formulation was not standardized to hyperforin, which has been suggested by some as an important active ingredient.<sup>31,32</sup> Hypericum may be most effective in less severe major depression (eg, HAM-D scores <20), but further study of this possibility needs to be conducted according to the standard diagnostic criteria.

From a methodological point of view, this study can be considered an example of the importance of including inactive and active comparators in trials testing the possible antidepressant

effects of medications. In fact, without a placebo, hypericum could easily have been considered as effective as sertraline, as some studies have done with respect to active antidepressants.<sup>2,3</sup> On the other hand, without sertraline as an active comparator, the results would have been interpreted as evidence for the lack of efficacy of hypericum, without consideration of the possibility that a low assay sensitivity of the trial might have contributed to the finding.

An increasing number of studies have failed to show a difference between active antidepressants and placebo.<sup>33,34</sup> Many of the presumed factors underlying this phenomenon were carefully attended to in this study, eg, adherence to quality control by rater training, treatment adherence monitoring, inclusion of experienced investigators, and carefully defined entry criteria. Despite all of this, sertraline failed to separate from placebo on the 2 primary outcome measures.

Besides the limitations already discussed, our study tested one particular hypericum extract, although many are marketed. Because the active ingredient of hypericum is unclear, it is difficult to extrapolate clinical data from one extract to other products. The extract we tested is among the best characterized, however, and is the one for which the most efficacy data are available. Thus, we believe that the results can be considered relevant to other hypericin-standardized hypericum extracts.

Because hypericum is widely available, it is likely to be used for milder depression, but its use in this population cannot be supported until trials show clear evidence of efficacy. According to available data, hypericum should not be substituted for standard clinical care of proven efficacy, including antidepressant medications and specific psychotherapies, for the treatment of major depression of moderate severity.

**Author Affiliations** *Coordinating Center, Duke University Medical Center, and Duke Clinical Research Institute.* Jonathan R. T. Davidson, MD, Kishore M. Gadde, MD, John A. Fairbank, PhD, K. Ranga Rama

Krishnan, MD, Robert M. Califf, MD, Cynthia B. Nanay, BSN.

*Research Triangle Institute.* Corette B. Parker, DrPH, Norma Pugh, MS, Tyler D. Hartwell, PhD.

*National Institute of Mental Health.* Benedetto Vi- tiello, MD, Louise Ritz, MBA, Joanne Severe, MS.

*Clinical Sites.* Jonathan O. Cole, MD (Harvard Medical School, McLean Hospital, Belmont, Mass); Charles de Battista, MD (Stanford University, Palo Alto, Calif); P. Murali Doraiswamy, MD (Duke University Medical Center, Durham, NC); John P. Feighner, MD (Feighner Research Institute, San Diego, Calif); Paul Keck, MD (University of Cincinnati, Cincinnati, Ohio); Jeffrey Kelsey, MD, PhD (Emory University Medical School, Atlanta, Ga); Khae-Ming Lin, MD (University of California, Los Angeles); Peter D. Lonnberg, MD (Summit Research Network, Seattle, Wash); Charles B. Nemeroff, MD, PhD (Emory University Medical School, Atlanta, Ga); Alan F. Schatzberg, MD (Stanford University, Palo Alto, Calif); David V. Sheehan, MD (University of South Florida, Tampa); Ram K. Srivastava, MD (Eastside Comprehensive Medical Services, New York, NY); Leslie Taylor, MD (Dean Foundation and University of Wisconsin, Madison); Madhukar H. Trivedi, MD (University of Texas Southwestern Medical Center, Dallas); Richard H. Weisler, MD (Duke University Medical Center, Durham, NC, and private practice, Raleigh, NC).

**Financial Disclosures:** Dr Davidson holds stock in Pfizer, American Home Products, GlaxoSmithKline, Procter and Gamble, and Triangle Pharmaceuticals; has received speaker fees from Solvay, Pfizer, GlaxoSmithKline, Wyeth-Ayerst, Lichtwer, and the American Psychiatric Association; has been a scientific advisor to Allergan, Solvay, Pfizer, GlaxoSmithKline, Forest Pharmaceuticals Inc, Eli Lilly, Ancile, Roche, Novartis, and Organon; has received research support from the National Institute of Mental Health (NIMH), Pfizer, Solvay, Eli Lilly, GlaxoSmithKline, Wyeth-Ayerst, Organon, Forest Pharmaceuticals Inc, PureWorld, Allergan, and Nutrition 21; has received drugs for studies from Eli Lilly, Schwabe, Nutrition 21, PureWorld Botanicals, and Pfizer; and has received royalties from MultiHealth Systems Inc, Guilford Publications, and the American Psychiatric Association. Dr Keck has been a consultant for and received grant support from Pfizer. Dr Kelsey has received research support from Abbott Laboratories, Bristol-Myers Squibb, Cyberonics, Eli Lilly, Forest Pharmaceuticals Inc, GlaxoSmithKline, Merck, Mitsubishi, Organon, Pfizer, Pharmacia-Upjohn, Sanofi, Solvay, and Wyeth-Ayerst; has been a member of the speakers bureau for Abbott Laboratories, Bristol-Myers Squibb, Forest Pharmaceuticals Inc, GlaxoSmithKline, Pfizer, Pharmacia-Upjohn, Solvay, and Wyeth-Ayerst; and has been a consultant for Eli Lilly, Skila Inc, and Wyeth-Ayerst. Dr Nemeroff has been a consultant for Abbott Laboratories, Acadia Pharmaceuticals, AstraZeneca, Bristol-Myers Squibb, Cypress Biosciences, Forest Laboratories, Janssen Pharmaceuticals, Eli Lilly, Merck, Neurocrine Biosciences, Organon, Otsuka, Pharmacia-Upjohn, Sanofi, GlaxoSmithKline, Somerset, Vela Pharmaceuticals, and Wyeth-Ayerst; has received grant support from Abbott Laboratories, AstraZeneca, Bristol-Myers Squibb, Forest Laboratories, Janssen Pharmaceuticals, Eli Lilly, National Alliance for Research on Schizophrenia and Depression, National Institute of Mental Health, Organon, Pfizer, Pharmacia-Upjohn, GlaxoSmithKline, Solvay, the Stanley Foundation and National Alliance of the Mentally Ill, and Wyeth-Ayerst; and has served on the speakers bureau of Abbott Laboratories, AstraZeneca, Bristol-Myers Squibb, Forest Laboratories, Janssen Pharmaceutica, Eli Lilly, Organon, Pfizer, Pharmacia-Upjohn, GlaxoSmithKline, and Wyeth-Ayerst. Dr Weisler has received grant support for, served on the speakers bureau of, consulted for, and owns stock in Pfizer. Dr Schatzberg has served as a consultant to Eli Lilly, GlaxoSmithKline, Wyeth-

Ayerst, Forest Laboratories, Organon, Solvay, and Pfizer. Dr Trivedi has received grant support from Abbott Laboratories, Akzo Nobel, Bayer, Bristol-Myers Squibb, Eli Lilly, Forest Laboratories, GlaxoSmithKline, Janssen Pharmaceutica, Johnson and Johnson, NIMH, Mead Johnson, Parke-Davis, Pfizer, Pharmacia-Upjohn, and Wyeth-Ayerst. Dr de Battista has served on the speakers bureau of Abbott Laboratories, Bristol-Myers Squibb, Cephalon, Eli Lilly, Forest Laboratories, GlaxoSmithKline, Organon, Pharmacia-Upjohn, Pfizer, and Wyeth-Ayerst; has received grants or research support from Bristol-Myers Squibb, Cephalon, Eli Lilly, NARSAD, NIMH, Nancy Pritzker Network, Pharmacia-Upjohn, GlaxoSmithKline, and Wyeth-Ayerst; has been a consultant for Bristol-Myers Squibb, Corcept Therapeutics Inc, Eli Lilly, GlaxoSmithKline, and Wyeth-Ayerst; and is a major stockholder in Corcept Therapeutics Inc. Dr Krishnan has received grant support from Eisai, Novartis, Pfizer, and Forest Laboratories and has served as a consultant to Abbott Laboratories, American Home Products, GlaxoSmithKline, Merck, Organon, Pfizer, Vela Pharmaceuticals, and Synaptic. Dr Doraiswamy has received research grants, honoraria, or both from Pfizer, Merck, GlaxoSmithKline, Johnson and Johnson, Wyeth-Ayerst, Organon, Forest Laboratories, and Eli Lilly and has received research support from NIMH.

**Author Contributions:** As primary investigator of the study, Dr Davidson had full access to all data and takes responsibility for the integrity of the data and accuracy of the data analysis.

**Study concept and design:** Davidson, Hartwell, Fairbank, Vitiello, Kelsey, Nemeroff, Weisler, Srivastava, Califf, Trivedi, Londborg, Krishnan, Gadde, Taylor, Cole, Doraiswamy.

**Acquisition of data:** Davidson, Parker, Fairbank, Ritz,

Keck, Kelsey, Nemeroff, Weisler, Binanay, Trivedi, de Battista, Londborg, Gadde, Feighner, Taylor, Cole, Doraiswamy, Sheehan, Lin.

**Analysis and interpretation of data:** Davidson, Pugh, Parker, Fairbank, Vitiello, Severe, Nemeroff, Schatzberg, Califf, Londborg, Krishnan, Gadde.

**Drafting of the manuscript:** Davidson, Parker, Vitiello, Severe, Nemeroff, Srivastava, Trivedi, de Battista, Gadde.

**Critical revision of the manuscript for important intellectual content:** Davidson, Hartwell, Pugh, Parker, Fairbank, Vitiello, Ritz, Keck, Kelsey, Nemeroff, Weisler, Srivastava, Schatzberg, Binanay, Califf, Trivedi, Londborg, Krishnan, Gadde, Feighner, Taylor, Cole, Doraiswamy, Sheehan, Lin.

**Statistical expertise:** Hartwell, Pugh, Parker.

**Obtained funding:** Davidson, Hartwell, Fairbank, Nemeroff, Califf.

**Administrative, technical, or material support:** Davidson, Parker, Fairbank, Vitiello, Ritz, Severe, Keck, Kelsey, Nemeroff, Weisler, Schatzberg, Binanay, Califf, Trivedi, de Battista, Londborg, Krishnan, Gadde, Taylor, Cole, Doraiswamy, Sheehan, Lin.

**Study supervision:** Vitiello, Weisler, Srivastava, Schatzberg, Binanay, Trivedi, Londborg, Gadde, Feighner, Cole, Lin.

**Funding/Support:** The research was supported by contract N01MH70007 from the National Center for Complementary and Alternative Medicine and the National Institute of Mental Health to Duke University Medical Center (Dr Davidson). Study medications were donated by Lichtwer Pharma and Pfizer.

**Acknowledgment:** We acknowledge the guidance of the following members of the study scientific advisory board: Joseph Betz, PhD (US Food and Drug Administration, Rockville, Md), Jacques Bradwejn, MD

(University of Ottawa, Ontario, Canada), Samuel Gershon, MD (University of Pittsburgh Medical Center, Pittsburgh, Pa), Robert Golden, MD (University of North Carolina, Chapel Hill), Jeffrey A. Lieberman, MD (University of North Carolina, Chapel Hill), Lydia Lewis (National Depressive and Manic-Depressive Association, Chicago, Ill), John E. Overall, PhD (University of Texas Health Sciences Center, Houston), and Richard Shader, MD (Tufts University, Boston, Mass). We are also indebted to Miriam Gibbon, MSW, and Janet Williams, DSW (Columbia University Biometrics Unit, New York, NY), Marguerite Evans, MS (National Center for Complementary and Alternative Medicine, Bethesda, Md), William Harlan, MD (National Institutes of Health, Bethesda, Md), Grayson Norquist, MD (National Institute of Mental Health, Bethesda, Md), Patricia Gibbons (National Institute of Mental Health), Geoffrey Chung, PhD (National Institutes of Health), Christine Swanson, PhD (National Institutes of Health Office of Dietary Supplements, Bethesda, Md), Deborah Roth, MSPH, Judith Kramer, MD, Rebecca Arias, BS, and Isa Lamerton, BA (Duke Clinical Research Institute, Durham, NC); Donna Medeiros, BS, Jim Wright, Jr, BA, and Deborah McFadden, MBA (Research Triangle Institute, Research Triangle Park, NC); and the National Institute of Mental Health data and safety monitoring board. The original protocol was designed by the National Institutes of Health Office of Complementary and Alternative Medicine and was subsequently adapted according to input from the investigators, scientific advisors, and data and safety monitoring board. Conduct of the study, analysis of data, and approval of the manuscript were undertaken with participation between the investigators, the National Institutes of Health, and the scientific advisors by means of conference calls and meetings.

## REFERENCES

- Linde K, Ramirez G, Mulrow CD, Pauls A, Weidenhammer W, Melchart D. St. John's wort for depression. *BMJ*. 1996;313:253-258.
- Wheatley D. LI-160, an extract of St. John's wort, versus amitriptyline in mildly to moderately depressed outpatients: a controlled 6-week clinical trial. *Pharmacopsychiatry*. 1997;30:77-80.
- Vorbach EU, Arnoldt KH, Hubner WE. Efficacy and tolerability of St. John's wort extract LI 160 versus imipramine in patients with severe depressive episodes according to ICD-10. *Pharmacopsychiatry*. 1997;30(suppl 2):81-85.
- Philipp M, Kohner R, Hiller KO. Hypericum extract versus imipramine or placebo in patients with moderate depression. *BMJ*. 1999;319:1534-1539.
- Woelk H, for the Remotiv/Imipramine Study Group. Comparison of St. John's wort and imipramine for treating depression. *BMJ*. 2000;321:536-539.
- Schrader E, for the Study Group. Equivalence of St. John's wort extract (Ze 117) and fluoxetine. *Int Clin Psychopharmacol*. 2000;15:61-68.
- Kalb R, Trautmann-Sponsel M, Kieser M. Efficacy and tolerability of hypericum extract WS 5772 versus placebo in mildly to moderately depressed patients. *Pharmacopsychiatry*. 2001;34:96-103.
- Montgomery SA, Hubner WD, Grigoleit HG. Efficacy and tolerability of St. John's wort extract compared to placebo in patients with mild to moderate depressive disorder. *Phytomedicine*. 2000;7(suppl 11):7.
- Shelton RC, Keller MB, Gelenberg AJ, et al. Effectiveness of St. John's wort in major depression. *JAMA*. 2001;285:1978-1986.
- Piscitelli SC, Burstein AH, Chait D, Alfaro R, Falloon J. Indinavir concentrations and St. John's wort. *Lancet*. 2000;355:547-548.
- Ruschitzka F, Meier PJ, Turina M, Luscher TF, Noll G. Acute heart transplant rejection due to Saint John's wort. *Lancet*. 2000;355:548-549.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
- First MB, Spitzer RL, Gibbon M, et al. *Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version, Modified for NIMH Clinical Trial of Hypericum in Depressive Disorder (SCID-Hypericum)*. New York, NY: Biometrics Research, New York Psychiatric Institute; 1997.
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23:56-62.
- Guy W, ed. *ECDEU Assessment Manual for Psychopharmacology*. Rockville, Md: National Institutes of Mental Health; 1976. Publication ADM 76-338.
- Beck AT, Steer RA, Brown GK. *The Beck Depression Inventory*. 2nd ed. San Antonio, Tex: Psychological Corp; 1993.
- Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. *Int Clin Psychopharmacol*. 1996;11(suppl):89-95.
- Brockmoller J, Reum T, Bauer S, Kerb R, Heber W-D, Roots I. Hypericin and pseudohypericin. *Pharmacopsychiatry*. 1997;30:102-107.
- Wagner H, Bladt S. Pharmaceutical quality of hypericum extracts. *J Geriatr Psychiatry Neurol*. 1994;7:565-568.
- Connor KM, Davidson JRT, Churchill LE. Adverse-effect profile of kava. *CNS Spectrums*. 2001;6:848-853.
- Bech P, Cialdella P, Haugh MC, et al. Meta-analysis of randomized controlled trials of fluoxetine v. placebo and tricyclic antidepressants in the short-term treatment of major depression. *Br J Psychiatry*. 2000;176:421-428.
- Moncrieff J, Wessley S, Hardy R. Meta-analysis of trials comparing antidepressants with active placebo. *Br J Psychiatry*. 1998;172:227-231.
- Fabre LF, Abuzzahab FS, Amin M, et al. Sertraline safety and efficacy in major depression. *Biol Psychiatry*. 1995;38:592-602.
- Lydiard RB, Stahl SM, Hertzman M, Harrison WM. A double-blind, placebo-controlled study comparing the effects of sertraline versus amitriptyline in the treatment of major depression. *J Clin Psychiatry*. 1997;58:484-491.
- Reimherr FW, Chouinard G, Cohn CK, et al. Antidepressant efficacy of sertraline. *J Clin Psychiatry*. 1990;51(suppl B, pt 12):18-27.
- Morris JB, Beck AT. The efficacy of antidepressant drugs. *Arch Gen Psychiatry*. 1974;30:667-676.
- Khan A, Warner HA, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials. *Arch Gen Psychiatry*. 2000;57:311-317.
- Schweizer E, Rynn M, Mandos LA, et al. The antidepressant effect of sertraline is not enhanced by dose titration. *Int Clin Psychopharmacol*. 2001;16:137-143.
- Suri RA, Altschuler LL, Rasgon NL, et al. Efficacy and response time to sertraline vs fluoxetine in the treatment of unipolar major depressive disorder. *J Clin Psychiatry*. 2000;61:942-946.
- Fabre LF, Abuzzahab FS, Amin M, et al. Sertraline safety and efficacy in major depression. *Biol Psychiatry*. 1995;38:592-602.
- Laakman G, Schule C, Baghai T, Kiser M. St. John's wort in mild to moderate depression. *Pharmacopsychiatry*. 1998;31(suppl 1):54-59.
- Wannemann M, Singer A, Muller WE. Inhibition of synaptosomal uptake of <sup>3</sup>H-L-glutamate and <sup>3</sup>H-GABA by hyperforin, a major constituent of St. John's wort. *Neuropharmacology*. 2000;23:188-197.
- Montgomery SA. The failure of placebo-controlled studies. *Eur Neuropharmacology*. 1999;9:271-276.
- Robinson DS, Rickels K. Concern about clinical drug trials. *J Clin Psychopharmacol*. 2000;20:593-596.