Effectiveness of St John’s Wort in Major Depression
A Randomized Controlled Trial

Richard C. Shelton, MD  
Martin B. Keller, MD  
Alan Gelenberg, MD  
David L. Dunner, MD  
Robert Hirschfeld, MD  
Michael E. Thase, MD  
James Russell, MD  
R. Bruce Lydiard, MD, PhD  
Paul Crits-Christoph, PhD  
Robert Gallop, PhD  
Linda Todd  
David Hellerstein, MD  
Paul Goodnick, MD  
Gabor Keitner, MD  
Stephen M. Stahl, MD  
Uriel Halbreich, MD

Context  Extracts of St. John’s wort are widely used to treat depression. Although more than 2 dozen clinical trials have been conducted with St John’s wort, most have significant flaws in design and do not enable meaningful interpretation.

Objective  To compare the efficacy and safety of a standardized extract of St John’s wort with placebo in outpatients with major depression.

Design and Setting  Randomized, double-blind, placebo-controlled clinical trial conducted between November 1998 and January 2000 in 11 academic medical centers in the United States.

Participants  Two hundred adult outpatients (mean age, 42.4 years; 67.0% female; 85.9% white) diagnosed as having major depression and having a baseline Hamilton Rating Scale for Depression (HAM-D) score of at least 20.

Intervention  Participants completed a 1-week, single-blind run-in of placebo, then were randomly assigned to receive either St John’s wort extract (n=98; 900 mg/d for 4 weeks, increased to 1200 mg/d in the absence of an adequate response thereafter) or placebo (n=102) for 8 weeks.

Main Outcome Measures  The primary outcome measure was rate of change on the HAM-D over the treatment period. Secondary measures included the Beck Depression Inventory (BDI), Hamilton Rating Scale for Anxiety (HAM-A), the Global Assessment of Function (GAF) scale, and the Clinical Global Impression–Severity and –Improvement scales (CGI-S and CGI-I).

Results  The random coefficient analyses for the HAM-D, HAM-A, CGI-S, and CGI-I all showed significant effects for time but not for treatment or time-by-treatment interaction (for HAM-D scores, \( P < .001 \), \( P = .16 \), and \( P = .58 \), respectively). Analysis of covariance showed nonsignificant effects for BDI and GAF scores. The proportion of participants achieving an a priori definition of response did not differ between groups. The number reaching remission of illness was significantly higher with St John’s wort than with placebo (\( P = .02 \)), but the rates were very low in the full intention-to-treat analysis (14/98 [14.3%] vs 5/102 [4.9%], respectively). St John’s wort was safe and well tolerated. Headache was the only adverse event that occurred with greater frequency with St John’s wort than with placebo (39/95 [41%] vs 25/100 [25%], respectively).

Conclusion  In this study, St John’s wort was not effective for treatment of major depression.


©2001 American Medical Association. All rights reserved.
MAJOR DEPRESSION AND ST JOHN’S WORT

for standardization, or of total extract varied from 0.4 to 5.4 mg and 300 to 1800 mg, respectively. In 13 studies comparing a single St John’s wort extract dose with placebo, 55.1% of St John’s wort–treated depressed patients showed significant improvement vs 22.3% with placebo. In comparison with standard antidepressants (imipramine, amitriptyline, or maprotiline) in 3 trials, 63.9% responded to St John’s wort compared with 58.5% with the antidepressant. Adverse effect problems with St John’s wort were low, especially compared with antidepressants. More common adverse effects included photosensitivity, headache, dry mouth, dizziness, constipation, and other gastrointestinal tract effects. However, only 4% of St John’s wort–treated patients vs 7.7% of those treated with antidepressants dropped out because of adverse effects. These results suggest that the St John’s wort extract may be a safe and effective alternative to antidepressants.

As noted by Linde et al2 and other recent reviews, most or perhaps all of the trials used in recent meta-analyses have had serious methodological flaws, thereby undermining confidence in their results (Table 1). The literature strategies used to uncover the studies listed in Table 1 included the use of a MEDLINE literature search (January 1967 to January 2001). In addition, the reference lists of relevant publications were reviewed. The problems in these studies included (1) not using standardized diagnostic practices, resulting in the inclusion of diagnostically heterogeneous groups; (2) failing to use standardized symptom rating instruments, such as the Hamilton Rating Scale for Depression (HAM-D), (3) relatively short duration of many studies; and (4) use of inexperienced investigators. Several trials compared St John’s wort vs an antidepressant but without a placebo control and often the sample sizes yielded insufficient statistical power to detect meaningful differences. In the absence of a placebo, a null result could be the equivalent of a failed trial, rather than support for the equal effectiveness of the experimental agent. Furthermore, the studies using an active drug for comparison were confounded by the use of inadequate doses of the antidepressant, typically amitriptyline or imipramine 100 mg/d or less. In no case were drug plasma levels monitored to ensure treatment adequacy. Finally, concerns have been raised that in these studies the blind may have been transparent because of the adverse effects of the comparison drug, a failure to adequately mask the taste of the St John’s wort product vs placebo, or other reasons.

Nonetheless, the cited reviews have concluded that St John’s wort extract appears to be a safe and effective alternative to antidepressants in the treatment of depression. For example, Gaster and Holroyd2 identified 8 controlled trials that they determined were of sufficient methodological rigor to support the effectiveness of St John’s wort in depression. However, elsewhere they note that there was “at least one significant methodological flaw in all but two of the studies.” In fact, there were problems with even these 2. In the first study,32 a comparison of St John’s wort vs amitriptyline, the dose of the tricyclic was only 75 mg/d, which is half or less of the dose typically required to produce a significant drug-placebo difference. Despite the inadequacy of the dosing, amitriptyline was significantly more effective than St John’s wort at the 6-week end point of the trial (P<.05). However, the authors concluded that St John’s wort was equal to amitriptyline.32 The second report32 was a comparison of St John’s wort vs placebo. In this study, the sample was mildly depressed (HAM-D 21-item scale score of 16-24), which typically would be expected to result in a higher placebo response rate,39-42 and the dose of St John’s wort extract was low (500 mg/d). They found that 56% of the St John’s wort–treated sample achieved the predetermined definition of response vs 15% of the placebo group. Not only did the placebo condition have a low response rate, but the average patient in this group experienced an 18% worsening in HAM-D symptom severity, a result that is highly unusual for a mildly depressed sample. Such a negative placebo response again raises some concern about the effectiveness of the blinding of the trial.

Considering how pervasive and serious the methodological problems of earlier studies of St John’s wort were, it is difficult to draw firm conclusions from them. Because of the widespread use of St John’s wort for the treatment of depression in the United States and elsewhere, we concluded that a carefully conducted, large-scale, placebo-controlled study of St John’s wort in depression was warranted. This trial was intended to determine the acute antidepressant efficacy, safety, and tolerability of a standardized extract of St John’s wort for the treatment of major depressive disorder. We tested the hypothesis that St John’s wort extract would produce a superior effect in contrast to placebo in the treatment of major depression. The efficacy of St John’s wort for patients with relatively less severe initial depression was also examined in a subanalysis.

METHODS

Study Design

This trial was conducted at 11 geographically diverse academic medical centers between November 1998 and January 2000. All patients provided written informed consent and the study was approved by the institutional review board for each participating institution. Participants for the study were physically healthy male or female outpatients, 18 years or older, diagnosed as having major depressive disorder, single episode or recurrent, without psychotic features according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), of at least 4 weeks’ duration. Participants had a score of at least 20 on the HAM-D (17-item scale) at baseline. Prospective participants with the following DSM-IV diagnoses were excluded: current cognitive disorder, posttraumatic stress disorder, eating disorder, or a substance use disorder in the...
last 6 months; panic disorder in the last year; or current or past history of bipolar disorder or any psychotic disorder, or borderline, antisocial, or schizotypal personality disorder. Anyone with a prior adequate trial of St John’s wort (at least 450 mg/d) for the treatment of depression or those who had taken St John’s wort for any reason in the last month were excluded. To reduce the potential for including a treatment nonresponsive sample, participants who had failed to respond to a trial of an antidepressant (fluoxetine hydrochloride, 20 mg/d, for at least 4 weeks or the equivalent) in the current episode or who had failed to respond to more than 1 adequate trial of antidepressant

Table 1. Design Limitations of Previous Controlled Trials of St John’s Wort

<table>
<thead>
<tr>
<th>Source, y</th>
<th>Control</th>
<th>Study Duration, wk</th>
<th>Diagnostic Practices/ Heterogeneity†</th>
<th>Less Experienced Investigators‡</th>
<th>No Standardized Symptom Ratings§</th>
<th>Low Depression Severity¶</th>
<th>Small Sample Size/ Inadequate Power‖</th>
<th>Low Comparator Dose/No Plasma Levels#</th>
<th>Low St John’s Wort Dose**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoffman and Kühl, 1979</td>
<td>Placebo</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steger, 1985</td>
<td>Desipramine</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warnecke, 1986</td>
<td>Diazepam</td>
<td>12</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schich et al, 1987</td>
<td>Placebo</td>
<td>4</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kniebel and Burchard, 1988</td>
<td>Amitriptyline</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schmidt et al, 1989</td>
<td>Placebo</td>
<td>4</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Werth, 1989</td>
<td>Imipramine</td>
<td>2</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kugler et al, 1990</td>
<td>Bromazepam</td>
<td>4</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harrer et al, 1991</td>
<td>Placebo</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halama, 1991</td>
<td>Placebo</td>
<td>4</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osterheider et al, 1992</td>
<td>Placebo</td>
<td>8</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reh et al, 1992</td>
<td>Placebo</td>
<td>8</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lehrl and Woelk, 1993</td>
<td>Placebo</td>
<td>4</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schmidt and Sommer, 1993</td>
<td>Placebo</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quandt et al, 1993</td>
<td>Placebo</td>
<td>4</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>König, 1993</td>
<td>Placebo</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td></td>
<td>NS</td>
</tr>
<tr>
<td>Bergmann et al, 1993</td>
<td>Amitriptyline</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hübner et al, 1994</td>
<td>Placebo</td>
<td>4</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harrer et al, 1994</td>
<td>Maprotiline</td>
<td>4</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hänsgen et al, 1994</td>
<td>Placebo</td>
<td>4</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sommer and Harrer, 1994</td>
<td>Placebo</td>
<td>4</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ditzler et al, 1994</td>
<td>Placebo</td>
<td>8</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td>NS</td>
</tr>
<tr>
<td>Vorbach et al, 1994</td>
<td>Imipramine</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Witte, 1995</td>
<td>Placebo</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheatley, 1997</td>
<td>Amitriptyline</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vorbach et al, 1997</td>
<td>Imipramine</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schrader et al, 1998</td>
<td>Placebo</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philipp et al, 1999</td>
<td>Placebo, imipramine</td>
<td>8</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brenner et al, 2000</td>
<td>Sertraline</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schrader, 2000</td>
<td>Fluoxetine</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woelk, 2000</td>
<td>Imipramine</td>
<td>6</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*NS indicates unclear or not specified.  
†Nonstandardized research diagnostic practices resulting in diagnostic heterogeneity.  
‡Use of investigators without apparent experience in psychiatry or research.  
§Standardized measurements of depressive symptoms were not used.  
‖Inclusion of mildly depressed subjects (eg, Hamilton Depression Scale score of <18).  
¶Statistical power was too low to detect meaningful differences between groups.  
#Dose of comparator drug was too low and/or no plasma levels were measured when available.  
**The extract dose in part or all of sample was less than 600 mg/d (standardized hypericin concentration).  
††Utilized infrequently used measures of depressive symptoms.  
‡‡Compared with a benzodiazepine (not an antidepressant).
in a previous episode also were excluded. Patients could not take other psychotropic medications during study participation, with the exception of zolpidem tartrate, which was allowed up to 10 mg/d for sleep for the first 3 weeks of the trial. All participants received a physical examination, electrocardiogram, hematological and blood chemistry screening, and urine testing for illicit drugs. Persons in psychotherapy were allowed if they had been in therapy for at least 3 months prior to baseline, and if the frequency of sessions did not change during participation. Women also received a urine pregnancy test. Participants were recruited from the participating clinics and by paid newspaper or radio advertisements.

The baseline assessment included the Structured Clinical Interview for DSM-IV Axis I Disorders.49 Qualified participants then began a single-blind placebo run-in period of 1 week. Those who experienced an improvement of greater than 25% or achieved a score of less than 20 on the 17-item HAM-D were excluded from further participation. The use of the placebo run-in is somewhat controversial.45,46 However, it remains a design feature that is used frequently in depression studies to minimize the impact of early placebo response. All others were randomly assigned in a double-blind fashion to receive either a standardized 300-mg tablet extract of St John's wort (Lichtwer Pharma GmbH, Berlin, Germany) or identically matched placebo for an 8-week treatment period. The St John's wort and placebo were prepared and distributed using procedures approved by the Food and Drug Administration under investigational new drug No. 56386. The preparation masked the taste and smell of St John's wort. The randomization schedule was determined independently at a site that prepared and packaged the St John's wort and placebo using a computer-generated random number program. The allocation schedule was maintained at this site and could be broken if needed in an emergency. Individual participants were assigned to a treatment condition according to the schedule after the completion of the run-in period. Individual lots of the St John's wort extract were evaluated for consistency by high-performance liquid chromatography with electrochemical detection using titrated hypericin as an internal standard. The intra-assay coefficient of variation was less than 5%. The dose of St John's wort or placebo was 1 tablet 3 times per day (equivalent to 900 mg/d of St John's wort) for at least 4 weeks. The dosage of St John's wort or placebo was increased to 4 tablets per day (equivalent to 1200 mg/d of St John's wort) for the remainder of the trial if there was insufficient improvement by week 4.

The use of placebo controls currently is highly controversial.47-49 Depression is a serious and potentially life-threatening condition. Therefore, extensive safeguards were needed. Participants were excluded if they posed a significant risk of suicide at any time during participation. Persons who scored greater than a 2 (ie, thoughts of death or wishes self dead, but no suicidal ideation or plan) on the suicide item of the HAM-D, or who were judged to have significant suicidal ideation or potential in the view of an investigator were excluded. Further, any clinically significant deterioration in the condition of the subject from baseline would result in exclusion. Those who left the study before completion were offered alternative, standard care immediately. All participants were provided antidepressant treatment for 6 months at no cost after participation.

Efficacy and safety assessments were performed at screening, baseline, and the end of weeks 1, 2, 4, 6, and 8. These evaluations included the HAM-D, the physician-rated Clinical Global Impression–Improvement (CGI-I), and –Severity (CGI-S) scales,50 vital signs, and a review of adverse events. The Beck Depression Inventory51 (BDI) (a self-rated scale) and the Global Assessment of Function scale were performed at screening, baseline, and week 8 (or end point). The Hamilton Anxiety (HAM-A) scale52 was performed at baseline and weeks 2 and 8. Laboratory assessments and an electrocardiogram were completed at screening and end point. Compliance was assessed by pill counts at each follow-up visit. Investigators performing Structured Clinical Interview for DSM-IV Axis I Disorder evaluations were certified by the Columbia University Research and Training Group and HAM-D evaluators by the Columbia University Rater Certification Group. The HAM-D assessments performed at baseline and end point were videotaped and reviewed by an independent assessor to ensure reliability.

**Statistical Methods**

Differences between the 2 treatment groups in baseline demographic and clinical characteristics were examined using independent t tests for continuous variables and by the Cochran-Mantel-Haenszel test for categorical variables, with site as the stratification variable. The primary efficacy analysis was a random coefficient regression model that examined differences in linear rate of change between St John's wort and placebo groups in HAM-D total scores over the course of the 8-week acute phase. This model measures each patient's deviation from the population average change over time and includes a random intercept and slope for each subject. For this model all available data (baseline and weeks 1, 2, 4, 6, and 8) were used. The model (performed using SAS version 6.12, PROC MIXED, SAS Institute Inc, Cary, NC) estimated fixed-effects for treatment and site. Population-averaged estimates for the linear trend over time and the linear trend over time per treatment are produced by this model (the latter testing the linear slope differences between treatment groups). The treatment-by-site interaction was not significant and was therefore dropped from the model. This evaluation was an intention-to-treat (ITT) analysis that included all patients who were randomized to treatment, even those patients without a postbaseline assessment. A secondary subgroup analysis was performed using only those patients with a baseline
HAM-D score at or below 22 (the median score in this sample).

Similar ITT mixed-effects analyses were performed on secondary efficacy variables (HAM-A, CGI-S, and CGI-I). Secondary analyses also examined HAM-D response and remission rates. Response was defined as a HAM-D score of 12 or less (representing at least a 50% improvement in HAM-D scores from baseline) and a CGI-I score of 1 or 2. Remission was defined as a HAM-D score of 7 or less with a CGI score of 1 or 2. In the full ITT sample, response and remission rates at week 8 were determined by a random coefficient regression analysis. Response and remission rates for the completer sample (those with a week 8 assessment) were also calculated using the week 8 HAM-D scores. Treatment differences in response and remission rates were examined using the Cochran-Mantel-Haenszel test (site as stratification variable).

For secondary variables assessed only at baseline and end point (ie, BDI, Global Assessment of Function scale), an analysis of covariance (ANCOVA) was performed using end point scores as the dependent variable, respective baseline scores as the covariate, and treatment group and site as main effects. A similar ANCOVA was also performed on HAM-D scores at end point. Treatment-by-site interactions were again not significant, and hence not included in the models.

Safety analyses comparing treatment groups on adverse events were conducted using the Fisher exact test. The proportions of discontinuations in each treatment group were compared with the Fisher exact test. All statistical tests were 2-tailed and significance was declared at the .05 level.

RESULTS

Baseline Clinical and Demographic Characteristics

Two hundred fifty-four patients were screened for the study, and 200 were randomized (Figure 1). Of the 54 who were not randomized, 34 (63.0%) did not meet study criteria. Sixteen were excluded because of improvement during the placebo run-in. The remaining patients were excluded for other reasons (eg, did not return, noncompliance, withdrew consent). Two hundred patients were randomized to either St John's wort (n=98) or placebo (n=102). Postbaseline follow-up data on the primary efficacy measure (HAM-D) was obtained on 95 (97%) of the 98 randomized to St John's wort and 100 (98%) of the 102 randomized to placebo. Although randomization was not stratified within site, there were no significant treatment group differences on any baseline demographic or clinical measure (Table 2), either across sites or within sites, with the exception of a difference in baseline HAM-A scores (placebo group had greater initial symptoms of anxiety). During the trial there were no significant protocol violations reported.

Relatively few additional comorbid psychiatric diagnoses were present. Three patients had a diagnosis of generalized anxiety disorder, 4 had a diagnosis of social phobia, and 7 had a lifetime history of dysthymia. Four patients in the St John's wort group and 8 patients in the placebo group were currently in psychotherapy at the time of randomization.

Study Treatment

For the modified ITT sample, the mean (SD) number of tablets taken daily was 3.7 (1.44) for the St John's wort group (n=91) (each tablet contained 300 mg) and 3.6 (1.32) for the placebo group (n=100). Those who completed 8 weeks of treatment were given a mean (SD) final daily dose of 3.9 (1.47) tablets (St John's wort: n=75) and 3.7 (1.35) tablets (placebo: n=86). Compliance was good and there were no differences between St John's wort or placebo groups.

Efficacy Analysis

The random coefficient regression model examining the rate of change in HAM-D scores over the 8 weeks of the trial for the ITT sample revealed a significant time effect (F1,194=172.4; P<.001), but no sig-
significant treatment effect ($F_{1,188}=2.0; P=.16$) or time-by-treatment interaction ($F_{1,193}=0.3; P=.58$). The slopes for the St John’s wort and placebo groups appear parallel (FIGURE 2). Similarly, estimated end point means from an ANCOVA, adjusting for baseline and site, were not statistically different ($F_{1,182}=0.7; P=.40$). Estimated mean for St John’s wort is 14.2 and estimated mean for placebo is 14.9 with a pooled standard deviation of 6.2.

In the ITT sample, there was not a significant difference (Cochran-Mantel-Haenszel $\chi^2=2.1; P=.15$) in week 8 response rates estimated from the random coefficient regression analysis (St John’s wort: 26/98 [26.5%; 95% confidence interval [CI], 18.7%-36.2%]; placebo: 19/102 [18.6%; 95% CI, 12.1%-27.5%]). A significantly higher remission rate was estimated for St John’s wort (14/98 [14.3%; 95% CI, 8.6%-22.8%]) compared with placebo (5/102 [4.9%; 95% CI, 2.0%-11.3%]) (Cochran-Mantel-Haenszel $\chi^2=5.2; P=.02$). However, this finding is likely influenced by the low variability around estimates that are close to zero (ie, remission rates were very low). In the completer sample, there was also a nonsignificant difference (Cochran-Mantel-Haenszel $\chi^2=3.3; P=.07$) in response rates between St John’s wort (26/79 [32.9%; 95% CI, 23.3%-44.1%]) and placebo (18/87 [20.7%; 95% CI, 13.3%-30.6%]). Similarly, the difference between St John’s wort (16/79 [20.3%; 95% CI, 12.7%-30.7%]) and placebo (9/87 [10.3%; 95% CI, 5.4%-18.8%]) in remission rates did not attain statistical significance (Cochran-Mantel-Haenszel $\chi^2=2.3; P=.07$).

Random coefficient regression analyses for the HAM-A total scores also evidenced a significant time effect ($F_{1,186}=72.0; P<.001$), and a significant treatment effect ($F_{1,186}=5.4; P=.02$), but failed to reveal a significant treatment-by-time interaction ($F_{1,186}=1.2; P=.27$). The significant treatment effect was a function of the previously mentioned difference between the groups in HAM-A scores at baseline. Similar results to the primary efficacy analysis of HAM-D scores were obtained with the CGI-S and CGI-I scales, with significant time effects ($F_{1,190}=100.9; P<.001; F_{1,194}=99.7; P<.001$, respectively), but no significant treatment effect ($F_{1,188}=1.0; P=.32; F_{1,188}=0.04; P=.84$) or treatment-by-time interactions ($F_{1,193}=0.3; P=.58; F_{1,191}=1.2; P=.28$). Mean scores, by visit, are given in TABLE 3. In the modified ITT sample, ANCOVAs yielded a nonsignificant effect for treatment on the BDI (1.140 =1.5; $P=.22$) and Global Assessment of Function scale ($F_{1,194}=2.3; P=.07$) (TABLE 3). Further, using a response score of less than 10 on the BDI, there were no differences between St John’s wort (24/75 [32%; 95% CI, 22.3%-43.5%]) and placebo (17/81 [21%; 95% CI, 13.3%-31.3%]) (Cochran-Mantel-Haenszel $\chi^2=2.3; P=.13$).

St John’s wort could be effective for less severely depressed patients. To address this, we divided the sample at the median HAM-D score of 22, with a score of less than 22 representing the less severely depressed group. There was no significant difference in rate of change in HAM-D scores between the St John’s wort group (n=60) and the placebo group (n=50) for those patients with relatively less severe initial depression by random coefficient analyses (treatment main effect: $F_{1,96}=0.2; P=.65$; treatment-by-time interaction: $F_{1,107}=0.2; P=.66$; slope for St John’s wort=−0.77 [95% CI, −0.98 to −0.57]; and slope for placebo=−0.71 [95% CI, −0.94 to −0.47]). Similarly, using estimated end point scores, there were no significant differences in responder rates (St John’s wort: 20/59 [33.9%; 95% CI, 22.8%-47.1%] and placebo: 11/50 [22.0%; 95% CI, 12.5%-35.8%]; Cochran-Mantel-Haenszel $\chi^2=1.0; P=.32$) or remission rates (St John’s wort: 12/59 [20.3%; 95% CI, 11.8%-32.8%] and placebo: 5/50 [10.0%; 95% CI, 4.1%-22.2%]; Cochran-Mantel-Haenszel $\chi^2=1.7; P=.20$) between treatment groups in the subgroup of patients with relatively milder initial levels of depressive symptoms.

| Table 2. Demographic and Clinical Characteristics of Randomized Patients at Baseline* |
|-----------------|-----------------|-----------------|
| Characteristic  | St John’s Wart  | Placebo         |
|                 | (n = 98)        | (n = 102)       |
| Women, %        | 64.9            | 62.8            |
| Age, mean (SD), y | 41.4 (2.5)      | 43.3 (3.7)      |
| White, %        | 85.6            | 86.1            |
| Marital status, % | Single          | 37.1            |
|                 | Married or cohabitating | 38.1 |
|                 | Widowed         | 2.1             |
| Widowed/separated | 22.7            | 25.7            |
| Depression diagnosis, % | Single episode | 31.6            |
|                 | Recurrent       | 66.3            |
|                 | Melancholic     | 39.0            |
| Age at onset of initial major depressive disorder, mean (SD), y | 29.9 (14.7) | 31.1 (16.2) |
| Duration of current major depressive disorder, mean (SD), y | 2.3 (6.3) | 2.7 (5.6) |
| Global Assessment of Function, mean (SD) | 56.3 (5.8) | 56.2 (4.9) |

*Approximately 80% of the volunteers were recruited by newspaper advertisements.

©2001 American Medical Association. All rights reserved.

Figure 2. Hamilton Rating Scale for Depression

Values are unadjusted mean scores.
Treatment Retention and Adverse Events

Attrition was similar across both treatment groups ($\chi^2_1=0.5; P=.50$) (Figure 1) with 80/98 (81.6%; 95% CI, 72.6%-88.2%) patients completing St John’s wort treatment and 87/102 (85.3%; 95% CI, 76.9%-91.0%) completing treatment in the placebo group. There were no discontinuations due to lack of efficacy. After randomization, 90/98 (5.1%) of participants receiving St John’s wort and 3/102 (2.9%) receiving placebo did not take the treatments and were excluded. Reasons for discontinuation after receiving at least 1 dose consisted of subject’s choice (2/98 [2.0%] for St John’s wort and 1/102 [1%] for placebo), lost to follow-up (8/98 [8.2%] for St John’s wort and 6/102 [5.9%] for placebo), noncompliance (0/98 [0%] for St John’s wort and 1/102 [1.0%] for placebo), and adverse events (1/98 [1.0%] for St John’s wort and 1/102 [1.0%] for placebo). For 8 patients, reason for early termination was not acquired (4/98 [4.1%] for St John’s wort and 4/102 [3.9%] for placebo).

Two adverse events (abdominal discomfort and headaches) occurred for 10% or more of the patients in 1 or both of the treatment groups. Of these, there was a significant difference ($P=.02$) between St John’s wort and placebo only for headaches, with a greater proportion of those taking St John’s wort reporting headaches (39/95 [41%; 95% CI, 30.5%-49.9%]) compared with placebo (25/100 [25%; 95% CI, 17.0%-33.9%]).

Based on the observed 18.6% response rate for placebo found in the current trial, a St John’s wort response rate of 36.1% would have been needed to achieve statistical power of 80% (2-tailed) using sample sizes of 100 per group. Although a smaller difference in response rates could obviously be detected with larger sample sizes, the current study was adequately powered to detect a clinically meaningful difference in response rates.

**COMMENT**

This study represents the first report of a large-scale, multicenter, randomized, placebo-controlled trial of St John’s wort extract in participants with major depressive disorder conducted in the United States. Unlike previous studies, St John’s wort extract failed to produce significant differences vs placebo on any of the outcome measures used: HAM-D, BDI, CGI, or HAM-A. Response rates in the ITT analysis were not significantly different (26.5% for St John’s wort vs 18.6% for placebo). The ITT sample showed significant differences in remission rates, but these rates were extremely low (14.3% for St John’s wort vs 4.9% for placebo). The response and remission rates also were not significantly different in those participants who completed the full 8-week trial. These results do not support significant antidepressant or anti-anxiety effects for St John’s wort when contrasted with placebo in a clinical sample of depressed patients.

St John’s wort was well tolerated by the sample, with better than 80% of both St John’s wort and placebo groups completing and 1% discontinuing due to adverse effects. Treatment-emergent adverse events that occurred in 10% or more of the sample included abdominal discomfort, insomnia, and headaches. Only the latter distinguished St John’s wort from placebo (40% vs 26%).

Given the extensive literature on the effectiveness of St John’s wort, an important question is why the data from the current report differ from previous outcome studies. Earlier research, including placebo and antidepressant comparison studies, have consistently shown superiority of St John’s wort when compared with placebo and equivalence to active drugs. However, as noted at the beginning of this article, these studies had significant flaws (Table 1). We attempted to address these important methodological issues in the design of this project. With nearly 30 controlled comparisons in depression, the lack of any published negative reports suggests systematic biases in evaluation or reporting. The results from the current study call into question the findings of the prior research reports.

It is important to appreciate that the homogeneity of the sample in this study imposes limitations on the generalizability of the findings, as is the case with most randomized controlled trials. The sample included outpatients with a baseline 17-item HAM-D score of greater than 20 and was collected from tertiary care clinics in academic medical centers. The average duration of depression in these participants was more than 2 years. In general, participants were not seeking treatment with herbal prod-

---

**Table 3. Secondary Efficacy Measures**

<table>
<thead>
<tr>
<th></th>
<th>Clinical Global Impression</th>
<th>Hamilton Anxiety Scale</th>
<th>Global Assessment of Function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beck Depression Inventory</td>
<td>CGI Severity</td>
<td>CGI Improvement</td>
</tr>
<tr>
<td>St John’s Wort</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>22.0 (8.2)</td>
<td>4.3 (0.5)</td>
<td>4.0 (0.7)</td>
</tr>
<tr>
<td>Week 1</td>
<td>16.0 (9.4)</td>
<td>3.3 (1.0)</td>
<td>2.9 (1.3)</td>
</tr>
<tr>
<td>Week 2</td>
<td>18.7 (9.9)</td>
<td>3.3 (1.0)</td>
<td>3.0 (1.1)</td>
</tr>
<tr>
<td>Week 4</td>
<td>16.0 (9.4)</td>
<td>3.3 (1.0)</td>
<td>3.0 (1.1)</td>
</tr>
<tr>
<td>Week 6</td>
<td>18.7 (9.9)</td>
<td>3.3 (1.0)</td>
<td>3.0 (1.1)</td>
</tr>
<tr>
<td>Week 8</td>
<td>16.0 (9.4)</td>
<td>3.3 (1.0)</td>
<td>3.0 (1.1)</td>
</tr>
</tbody>
</table>

*Values are expressed as mean (SD).
MAJOR DEPRESSION AND ST JOHN’S WORT

utcts. Therefore, patients who were more mildly and less chronically ill, those from other health care environments (eg, primary care practice), or persons who preferred alternative medicine approaches to their treatment might have a different outcome. We would encourage studies with different populations. However, the results of this study suggest that persons with significant major depression should not be treated with St John’s wort, given the morbidity and mortality risks of untreated or ineffectively treated major depression.53,54 Further, recent reports suggest a potential for interactions with certain drugs.55 Until well-designed positive studies are published, we conclude that there currently is no credible evidence to support the efficacy of St John’s wort for people with major depression.

Whence, then, St John’s wort? The data from the current study, generally, were negative. However, there are some hints that there could be at least some effect. For example, although the primary data analyses were negative, St John’s wort did produce a significantly greater proportion of remission in the ITT analysis compared with placebo. The response and remission rates for the 2 groups showed fairly consistent differences in both ITT and completer groups (about 8%-12%). More methodologically rigorous studies of depression are indicated. In addition, studies in other diagnostic groups, such as persons with primary anxiety disorders, may be warranted. However, claims for effectiveness for these indications should not be made until there is strong supporting data from well-controlled clinical trials.56-58

REFERENCES


The sea lies all about us. The commerce of all lands must cross it. The very winds that move over the lands have been cradled on its broad expanse and seek ever to return to it. The continents themselves dissolve and pass to the sea, in grain after grain of eroded land. So the rains that rose from it return again in rivers. In its mysterious past it encompasses all the dim origins of life and receives in the end, after, it may be, many transformations, the dead husks of that same life. For all at last returns to the sea—to Oceanus, the ocean river, like the ever-flowing stream of time, the beginning and the end.

—Rachel Carson (1907-1964)
6.1) increased risk of pancreatic cancer compared with nonsmoking members of the cohort. An unexpected finding was that pancreatic cancer developed 20 years earlier in smokers than in nonsmokers ($P = .02$). (FIGURE)

Forty-three percent of adults in the cohort had consumed alcohol, and drinking status was known for 17 of the 19 patients with pancreatic cancer. The mean (SD) age of onset of pancreatic cancer in 11 alcohol consumers was 55 (14.7) years, which was nearly identical to the mean age of onset of pancreatic cancer in 6 nonconsumers (56 [17.8] years). In multivariate analysis, a nonsignificant increased risk of pancreatic cancer was observed in consumers vs nonconsumers of alcohol (OR, 2.1; 95% CI, 0.7-6.3).

Comment. Hereditary pancreatitis usually begins in childhood and is now known to be caused by defects in the trypsinogen gene located at 7q35. Although the trypsinogen gene is not considered to be a cancer gene, the high risk of pancreatic cancer in these patients may be related to progressive glandular destruction over a long time period.

Our data suggest that in this sample of patients with hereditary pancreatitis, as in the general population, smoking doubles the risk of pancreatic cancer and accounts for approximately 25% to 30% of all pancreatic tumors. In addition, smoking appears to result in a dramatically earlier age of diagnosis of pancreatic cancer, perhaps because of a gene-environment interaction. We believe it is unlikely that either selective overreporting of smoking in young patients or underreporting of smoking in older patients with pancreatic cancer occurred. Evidence has suggested that there is a smoking-genetic interaction in patients with familial pancreatic cancer, but additional studies will be required to learn whether smoking has a similar effect in other patients who have an inherited susceptibility to pancreatic cancer. Meanwhile, our data indicate that in this germ-line disorder with a high risk of pancreatic cancer, smoking acts as an independent risk factor and seems to lower the age of onset of this aggressive cancer by approximately 2 decades. In addition, since alcohol can be injurious to the pancreas, we suggest that persons with a predisposition to pancreatic cancer should avoid drinking alcohol.

Albert B. Lowenfels, MD
Department of Surgery and Community and Preventive Medicine
New York Medical College
Valhalla, NY
Patrick Maisonneuve
Division of Epidemiology and Biostatistics
European Institute of Oncology
Milan, Italy
David C. Whitcomb, MD
Division of Gastroenterology, Hepatology, and Nutrition
Department of Medicine and Center for Genomic Sciences
University of Pittsburgh
Pittsburgh, Pa
Markus M. Lerch, MD
Department of Medicine B
Westfälische Wilhelms Universität
Münster, Germany
Eugene P. DiMagno, MD
Gastroenterology Research Unit
Mayo Clinic
Rochester, Minn

Funding/Support: This study was supported by grants from the C.D. Smithers Foundation; Solvay Pharmaceuticals Inc; the Italian Association for Cancer Research; the National Institute of Health (grant DK54709); the German Research Council, and the Interdisciplinary Center for Clinical Research (IZKF), Münster.


CORRECTION

Incorrect Spelling: In the Original Contribution entitled “Effectiveness of St John’s Wort in Major Depression: A Randomized Controlled Trial” published in the April 18, 2001, issue of THE JOURNAL (2001;285:1978-1986), an author’s name was misspelled. On page 1978, Paul Crites-Cristoph, PhD, should have been Paul Crites-Cristoph, PhD. On page 1985 in the Author Affiliations, his affiliation should have been “Department of Psychiatry, University of Pennsylvania, Philadelphia (Dr Crites-Cristoph and Gallop).” Also on page 1985, in the Author Contributions, Crites-Cristoph should appear after the headings “Study Concept and Design, Acquisition of Data, Analysis and Interpretation of Data, Drafting of the Manuscript, Critical Revision of the Manuscript for Important Intellectual Content, Statistical Expertise, Administrative, Technical, or Material Support, and Study Supervision.”

©2001 American Medical Association. All rights reserved.