Eur Arch Psychiatry Clin Neurosci (2005) 255 : 40–47 DOI 10.1007/s00406-004-0532-z

ORIGINAL PAPER

Lars Bjerkenstedt · Gunnar V. Edman · R. Gisberth Alken · Marcus Mannel

Hypericum extract LI 160 and fluoxetine in mild to moderate depression

A randomized, placebo-controlled multi-center study in outpatients

Received: 2 January 2004 / Accepted: 5 May 2004 / Published online: 12 November 2004

Abstract Efficacy and tolerability of *Hypericum* LI 160 was compared to fluoxetine and placebo in mild to moderate Major Depression (DSM-IV) in a 4-week randomized, double-blind trial. One hundred and sixty-three outpatients from 15 general practitioner centers received either 900 mg *Hypericum* LI 160, 20 mg fluoxetine, or placebo daily. Amelioration was measured by the Hamilton and the Montgomery-Åsberg Depression scales. Response and remission rates and global ratings by investigators and patients were measured. Adverse event reports, laboratory screening, vital signs, physical exams and ECG were collected. No significant differences could be observed regarding efficacy measures except for remission rate (*Hypericum* 24 %; fluoxetine 28 %; placebo 7 %). *Hypericum* was significantly better tolerated than fluoxetine. *Hypericum* LI 160 or fluoxetine were not more effective in short-term treatment in mild to moderate depression than placebo.

Key words *Hypericum* · fluoxetine · placebo · major depression · outpatient treatment

L. Bjerkenstedt, MD (✉)
Dept. of Clinical Neuroscience
Psychiatry Section
Karolinska Hospital
171 76 Stockholm, Sweden
Tel.: +46-8/51772901
E-Mail: lars.bjerkenstedt@nvso.sll.se

G. V. Edman
Dept. of Psychiatry
Research and Development Unit
Danderyd's Hospital
Danderyd, Sweden

R. G. Alken
Cooperative Clinical Drug Research and Development AG
Neuenhagen, Germany

M. Mannel
Lichtwer Pharma GmbH
Berlin, Germany

EAPCN 532

Introduction

The antidepressive effect of *Hypericum* extract has been recognized since medieval ages and is nowadays widely used as an herbal remedy. Scientific clinical investigations started in the late 1970s with a report of a placebo-controlled trial comprising 60 depressive patients (Hoffmann and Kühl 1979). Since then more than 30 controlled trials investigating standardized methanolic or ethanolic dry extracts from *Hypericum perforatum* L. in depressive disorders have been conducted, summarized in several reviews and meta-analyses (Gupta and Möller 2003; Kim et al. 1999; Linde et al. 1996). In 10 controlled trials, *Hypericum* extract LI 160 was shown to be superior to placebo or similarly effective as tricyclic antidepressants in the treatment of mild to moderate depressive episodes, and to have a more favorable safety profile (Volz 1997). In a review by Volz and Laux, the efficacy of St. John's wort (17 controlled studies) was compared with fluoxetine (nine controlled studies; Volz and Laux 2000). They found no significant differences in efficacy between the drugs. A careful examination of these studies showed no differences in number of adverse events between St. John's wort and placebo while fluoxetine regularly showed tolerability advantages towards tricyclic antidepressants. In this review, however, none of the studies conducted a direct comparison of the two drugs. The conclusion of equal efficacy of *Hypericum* extract and fluoxetine, and better tolerability of St. John's wort extract was confirmed by two recent, randomized, double blind studies (Harrer et al. 1999; Schrader 2000). Similar results were obtained in a study comparing *Hypericum* LI 160 with sertraline by Brenner and colleagues (Brenner et al. 2000). Of interest is a report by Vorbach and colleagues who found an antidepressive effect of St. John's wort (LI 160; 1200–1800 mg) in elderly patients suffering from *severe* major depression (Vorbach et al. 2000).

A major weakness of most of the efficacy studies of St. John's wort is the lack of placebo-control. A concept

paper on the revision of the European guideline on antidepressant medicinal products emphasizes the introduction of a placebo group in short-term efficacy trials of antidepressants (The European Agency for the Evaluation of Medicinal Products 1998). In the treatment of mild to moderate depression, the placebo response seems to be greater than in studies of severe depression (Khan et al. 2002). Furthermore, studies in mild to moderate depression with *hypericum* extracts should also have a standard antidepressant in a third arm serving as internal validation of the efficacy (Deltito and Beyer 1998). Since most of the patients fulfilling the diagnostic criteria for mild to moderate depression are treated in general practice (GP) as outpatients, tolerability is of great importance (Brenner et al. 2002). Thus, in case of equal efficacy, safety reasons should have a great impact on the choice of drug therapy. The aim of the present study was to investigate the efficacy and tolerability of LI 160, a St. John's wort dry extract, in mild to moderate depression as compared to fluoxetine and placebo in a prospective, randomized, double-blind, double-dummy parallel cohort trial.

Material and methods

Patients

Outpatients meeting the DSM-IV criteria for an acute, recurrent episode of Major Depression Disorder (MDD) with mild or moderate intensity (296.31, 296.32; American Psychiatric Association 1994) were recruited from general practitioners (GPs). Other inclusion criteria were: Caucasian females and males, age 18 to 70 years; a minimum total score of 21 on the 21-item Hamilton Depression Scale (HamD; Hamilton 1967); history of at least two episodes of non-psychotic MDD; capacity and willingness to give informed consent and to comply with study procedures.

Exclusion criteria were: a diagnosis of psychotic mental disorder; other disorders requiring concomitant psychoactive medication; MAO-I treatment within 14 days prior to entry; history of treatment-resistant MDD (at least two different antidepressants over 6 weeks at sufficient doses) from at least two previous depressive episodes; risk of suicide; history of seizure disorder; alcohol or substance abuse; other serious unstable acute or chronic medical illness; severely impaired hepatic or renal function; pregnancy, breast-feeding, or use of inadequate contraceptives in fertile women; known intolerance or hypersensitivity to study medications; substantial placebo response (HamD reduction ≥ 20 %) at the end of placebo run-in phase; treatment with any investigational drug during three months prior to inclusion; participation in another clinical trial within 30 days before start of the study.

The study was approved by Ethics Review Committees (ERC) of the Karolinska Institutet, Stockholm, the University of Lund and the University of Gothenburg, all ERCs in Sweden. The study was performed in accordance with Good Clinical Practice (GCP) guidelines and with the Declaration of Helsinki and its revisions (World Medical Association 2000). All patients gave written informed consent.

Between August 1997 and March 1999, a total of 15 GPs in the south of Sweden recruited patients suffering from mild to moderate major depression. Of 174 randomized subjects who passed the 3 to 7 days open placebo run-in phase, 170 individuals still meeting inclusion criteria entered the double-blind treatment phase. Since seven patients dropped out prior to the first follow-up visit, the intention-to-treat (ITT) sample comprised 163 subjects. One hundred and forty-eight patients completed the six-week treatment.

The proportion of patients diagnosed to suffer from mild/moderate depression (DSM-IV 296.31/296.32) was approximately 1:2 in all groups (*Hypericum*: n = 18/39; fluoxetine: n = 23/33; placebo: n = 19/38). Frequencies and types of concomitant diseases and medications (e. g. psychotropics such as oxazepam and zopiclone) were evenly distributed between treatment groups.

No statistically significant differences between the three treatment groups could be observed with respect to baseline demographics or clinical characteristics (Table 1).

A considerable number of patients taking concomitant psychotropic medications or lacking a sufficient washout period of pre-existing antidepressant medication (< 1 week) were excluded from the according-to-protocol (ATP) analysis. Therefore the ATP sample consisted of 115 patients. A detailed flow chart is presented in Fig. 1.

Study design

The study was a prospective, randomized, double-blind, placebo-controlled, parallel-group, multi-center, outpatient trial. Fluoxetine served as an active control. Patients were randomly allocated to receive one of the three treatments after a 3–7 day open placebo run-in phase.

Following inclusion, patients were seen after weeks 2 and 4. To fulfill a requirement of the ERC, placebo patients were randomly switched after 4 weeks to either *Hypericum* LI 160 or fluoxetine for another 14 days (until week 6, final visit) in a double-blind fashion.

Table 1 Baseline demographics and clinical characteristics of patients, showing absolute values, means ± SD and medians

Measure[a]	*Hypericum* LI 160	Fluoxetine	Placebo	
Intention-to-treat population	(n = 54)	(n = 54)	(n = 55)	p-value[b]
Sex	43f/11m	41f/13m	45f/10m	0.87
Age (years)	49.1 ± 12.0	50.4 ± 11.6	51.4 ± 11.8	0.58
Height (cm)	168.4 ± 9.6	167.7 ± 8.9	167.7 ± 7.9	0.89
Weight (kg)	75.7 ± 17.0	72.9 ± 14.9	70.7 ± 11.7	0.21
HamD (total score)	24.9 ± 4.8	23.8 ± 3.6	25.2 ± 4.4	NS
MADRS (total score)	25.5 ± 6.5	24.6 ± 6.2	25.4 ± 6.8	NS
Safety population	(n = 57)	(n = 56)	(n = 57)	p-value[c]
Duration of illness (years)	10.0	9.0	9.4	0.94
Duration of current episode (days)	85.0	76.0	86.0	0.62

[a] *HAMD* Hamilton Depression Scale; *MADRS* Montgomery Åsberg Depression Rating Scale; [b] P-values of ANOVA, except for categorical measures (Fisher's exact test); NS, not significant (p-value not shown due to complex test strategy given in the statistical plan; [c] P-values of Kruskal-Wallis Test

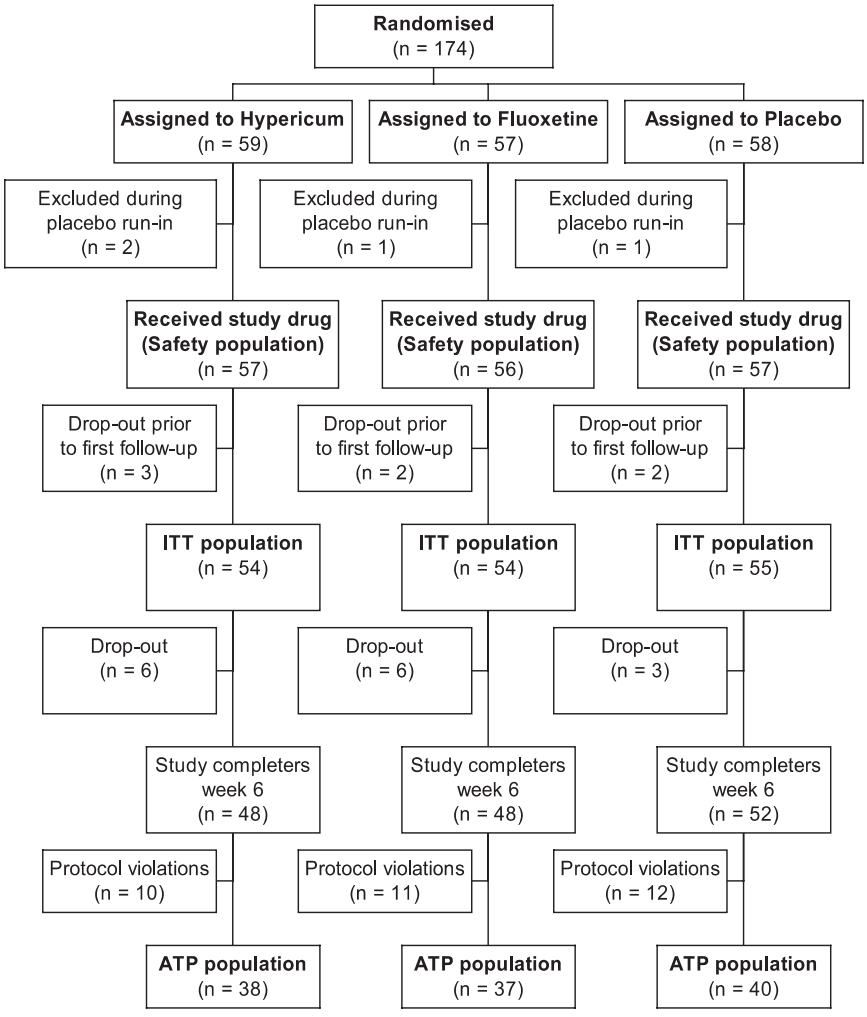

Fig. 1 Flow diagram of subject progress through the trial and allocation to analysis samples. *ITT* Intention-to-treat (comprising all patients with at least one HamD rating under treatment). *ATP* According-to-protocol (comprising all patients who have completed the trial and are lacking major protocol violations)

Thus, the primary end point for analysis was the end of week 4. To accomplish requirements from the ERCs, patients not exhibiting significant improvement of depressive symptoms during the 4-week double-blind treatment phase should be switched to open fluoxetine treatment and treated as dropouts. However, no such patient appeared. At the final visit after another 2 weeks, only patients treated with fluoxetine and *Hypericum* extract from the start were compared.

Depressive symptoms, adverse events, drug accountability (pill counting) and exclusion criteria were assessed at all visits from week 0 through week 6. At week 6, global judgments of efficacy and tolerability by investigator and patient were recorded. Physical exams, blood chemistry, hematology, ECG and urinalysis were done at screening and at week 6.

Study medications

All medications were provided by Lichtwer Pharma GmbH (Berlin, Germany) in accordance with GMP requirements.

At baseline, subjects received one of the following medications: one sugar coated tablet containing 300 mg *Hypericum* extract LI 160 (extraction solvent 80 percent methanol in water, drug-extract ratio 4–7:1; Lichtwer Pharma GmbH, Berlin Germany) t. i. d.; one hard gelatin capsule containing 20 mg fluoxetine-hydrochloride (Eli Lilly Sweden AB) in the morning; matching placebo tablets and capsules. A double-dummy technique with matching placebos for each active treatment was applied. Thus, both placebos were identical in shape, weight, color, smell, and taste to their corresponding verum formulations. Each participant received three sugar coated tablets and one capsule daily containing active substance or placebo.

Study drug was labeled sequentially by the manufacturing department of the sponsor on the basis of the randomization list, and was provided to the investigators in blocks of six. All investigators and personnel, actively involved in the trial, were blinded to group assignment until the database was closed. The investigator confirmed the reception of the study medication, including all follow-up supplies. After completion of the study, drug supplies not used were returned to the monitor.

Efficacy measures

Primary efficacy end-point was change in HamD total score from baseline to week 4 (*Hypericum* vs. placebo) and from baseline to week 6 (*Hypericum* vs. fluoxetine), respectively.

Secondary end points comprised treatment response at week 4 (decrease of more than 50 % of the HamD total score relative to baseline), remission (HamD total score < 8 at week 4), the HamD absolute values, the Montgomery Åsberg Depression Rating Scale (MADRS; Montgomery and Åsberg 1979) absolute values and pre-post differences, Clinical Global Impressions (CGI; Guy 1976), and global ratings of efficacy ("Excellent", "Moderate", "Slight", or "Ineffective") by patients and investigators (week 6). All clinical ratings by the investigators were performed after a training session, in which they were educated and trained in rating of depressive symptoms according to HamD and MADRS.

Safety measures

Based on a standardized patient interview with open questions, adverse events were recorded at each visit by the investigator, and classified according to the World Health Organization Adverse Reaction Terminology (WHO-ART; WHO 1999). Patient and investigator independently also rated tolerability on a 4-point scale ("Excellent", "Moderate", "Slight", or "Intolerable") at the end of therapy. Laboratory results, ECG recordings and physical examinations provided further information about safety and tolerability of treatments.

Statistics

The primary analysis included all randomized patients who had at least one follow-up visit under medication. Missing values during the 4-week treatment were handled conservatively by applying the last-observation-carried-forward (LOCF) method. At six weeks, only patients in the fluoxetine and *Hypericum* extract groups were compared, as the placebo group was shifted over to either fluoxetine or *Hypericum* extract. All patients who completed the 4-week treatment phase without major protocol violations were analyzed according-to-protocol.

Assessment of safety variables was carried out in the safety sample, comprising all randomized patients who took study drug at least once. Since placebo patients were randomly switched to one of the active treatments after 4 weeks, the reporting of safety data was focused on the first 4 weeks of treatment to ensure a fair evaluation.

Standard descriptive statistics (mean, standard deviations, and frequencies) were used to summarize characteristics of the samples. Tests for treatment differences included χ^2-test and Fisher's exact test for categorical and Student's t-test, ANOVA, Wilcoxon-Mann-Whitney-U test and Kruskal-Wallis test for continuous variables. All statistical tests were two-tailed with α set to 0.05.

Results

Efficacy

During the four weeks double-blind treatment, a 35–40 % reduction of HamD total score could be observed in the ITT sample without any significant differences between treatments (Fig. 2). After six weeks a further decrease in HamD total score to approximately 48 % was seen for both *Hypericum* and fluoxetine. This decrease was, however, not placebo-controlled. Almost identical results were attained with the MADRS also without any significant differences between groups.

The sadness items of the HamD (item 1) and of the MADRS (items 1 and 2) were analyzed separately. None of the comparisons of amelioration attained statistical significance (all F ratios for the interaction effects were below 1.00).

Moreover, no statistically significant differences between the groups regarding improvement were observed by means of the CGI-I. Of the *Hypericum* patients, 46 % were "much" or "very much" improved at the end of 4-week treatment compared to 41 % of the fluoxetine and 40 percent of the placebo group. After 6 weeks, the proportions were 57 % for *Hypericum* and 50 % fluoxetine.

The proportions of patients responding to treatment (HamD reduction > 50 % of baseline scores) were similar in all groups (*Hypericum* 38 %, fluoxetine 37 %, placebo 41 %). Applying the stricter criterion of remission (final HamD < 8) to assess treatment success, both *Hypericum* (24 %) and fluoxetine (28 %) were significantly superior to placebo (7 %; cf. Fig. 3).

Global judgments of efficacy by patients and investigators yielded no significant differences between treatments. Results from efficacy analyses in the ATP sample closely corresponded with ITT results. Major efficacy results are given in Table 2.

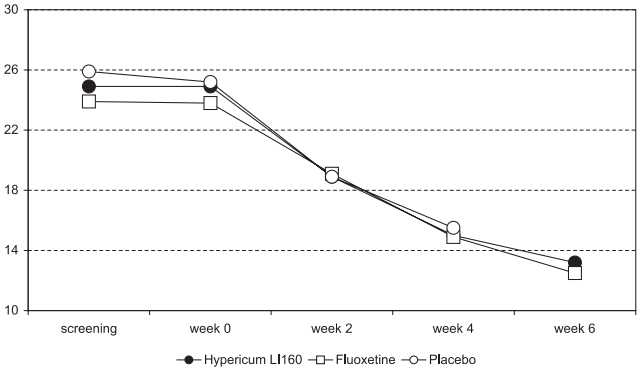

Fig. 2 Mean Hamilton Depression Scale total scores by week and treatment (ITT population, LOCF). Week 6 values of the placebo group are omitted, since placebo patients were switched after 4 weeks to either *Hypericum* or fluoxetine treatment. Week 6 values for *Hypericum* and fluoxetine do not include switched placebo subjects

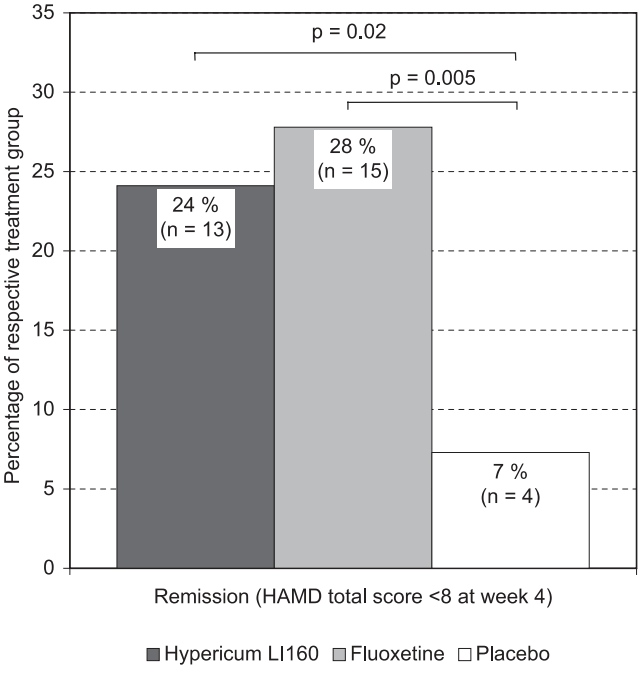

Fig. 3 Proportions of patients remitted (i. e. HamD total score below 8 at week 4) in the different treatment groups. P values represent two-sided probabilities of Fisher's exact test

Table 2 Summary of major efficacy results (ITT population)

Measure[a]	Hypericum LI 160 (n = 54)	Fluoxetine (n = 54)	Placebo (n = 55)	p-value[b]
HamD total score, baseline	24.9±4.5	23.8±3.7	25.2±4.6	0.26
HamD total score, week 4	15.0±8.4	14.9±8.4	15.5±6.7	0.70
Δ-HamD (week 4 – baseline)	9.9±8.1	8.9±8.0	9.7±7.0	0.90
MADRS total score, baseline	25.5±6.4	24.9±3.9	25.4±4.6	0.81
MADRS total score, week 4	15.6±10.4	16.5±9.9	17.1±8.7	0.29
Δ-MADRS (week 4 – baseline)	9.9±9.1	8.4±8.9	8.3±7.9	0.34
Response (HamD reduction week 4 > 50 %)	40.7 %	37 %	38.2 %	0.85
Remission (HamD total score week 4 < 8)	24.1 %	27.8 %	7.3 %	0.02

[a] *HamD* Hamilton Depression Scale; *MADRS* Montgomery Åsberg Depression Rating Scale; [b] p-values of statistical tests Hypericum vs. Placebo (Student's t-test for continuous variables; Fisher's exact test for categorical variables)

Safety

The mean duration of exposure during the 4-week double-blind treatment phase was 29.4, 28.9 and 28.9 days for *Hypericum*, fluoxetine, and placebo, respectively. One hundred and seventeen adverse events (AEs) were reported by 69 patients (Table 3). Both the number of AEs and the incidence of patients reporting AEs were higher in the fluoxetine group compared to *Hypericum* and placebo patients, the latter comparisons being statistically significant (p = 0.04 fluoxetine vs. *Hypericum*; p = 0.01 fluoxetine vs. placebo; Fisher's exact test). The incidence of patients reporting AEs was not significantly different between the *Hypericum* and the placebo group (p = 0.84; Fisher's exact test). A similar pattern, although not significant, could be observed regarding the assessment of the causal relationship between AE and study drug. Ten (n = 4 in *Hypericum*, n = 4 fluoxetine, and n = 2 placebo) subjects withdrew from the study because of adverse events. There were no significant differences in laboratory assessments, physical examinations, vital signs, or ECGs.

Discussion

Efficacy

In this study, mild to moderately depressed patients were randomly assigned to a treatment with either *Hypericum* (LI 160), fluoxetine, or placebo. Both active treatments failed to prove superiority over placebo regarding the primary efficacy measure, the HamD total score reduction from baseline to 4 weeks of treatment. The proportion of responders was almost the same for all groups and additional evaluations of the secondary variables comprising MADRS, CGI and global ratings revealed no significant differences.

The only variable that showed a better efficacy for *Hypericum* LI 160 and fluoxetine as compared to placebo was the percentage of patients who where remitted, defined as a HamD score below 8 at the end of treatment. Remission rates around 25 % for the two active drugs may appear to be low. However, taking into account that the assessment of efficacy was made after

Table 3 Frequencies of adverse events during the 4-week double-blind treatment phase (safety population)

	Hypericum LI 160 (n = 57)	Fluoxetine (n = 56)	Placebo (n = 57)	Total (n = 170)
No. of adverse events	38	52	27	117
No. of patients reporting adverse events (%)[1]	20 (35.1 %)	31 (55.4 %)	18 (31.6 %)	69 (40.6 %)
No. of adverse events rated to be "definitely", "probably" or "possibly" related to Study Drug	24	39	15	78
Frequencies of specific adverse events[2]				
Body as a whole	13	18	5	36
Gastro-intestinal system disorders	6	17	11	34
Autonomic nervous system disorders	10	12	8	30
Central & peripheral nervous system disorders	10	3	4	17
Skin and appendages disorders	9	5	3	17
Psychiatric disorders	2	8	3	13
Metabolic and nutritional disorders	–	6	–	6
Others	5	8	3	16
Total	55	77	37	169

[1] Hypericum vs. Fluoxetine: p = 0.04; Hypericum vs. Placebo: p = 0.84; Fluoxetine vs. Placebo: p = 0.01 (Fisher's exact test)
[2] Classification according to WHO-ART system organ classes; multiple entries of single adverse events were possible when a preferred term is allocated to more than one system organ class

only four weeks of treatment, this figure reflects a reasonable effect size (> 1.00) as compared to other depression trials.

Despite the DSM-IV classification into mild and moderate depression, at inclusion, 33 patients could be regarded as severely depressed according to a cut-off point of 31 in the MADRS (Müller et al. 2003). These patients were evenly distributed between treatments (12 or 22 % in the *Hypericum*, seven or 13 % in the fluoxetine, and 14 or 26 % in the placebo group) and cannot explain the lack of significant difference in efficacy.

All the raters were GPs and their ability of rating depressive symptoms might be questioned despite the training procedure. However, standard deviations (SD) for HamD and MADRS total score were not higher than expected (approximately 4.0–4.5 at week 2 for HamD and 6.5 for MADRS), which otherwise might have cast doubt on the skill of the raters. As a matter of fact, the SD was lower than that applied in the power analysis (SD = 7.0). Moreover, in Sweden as in most countries, mild to moderately depressed patients are treated by GPs. Thus, GPs are both educated to diagnose a patient as mild to moderately depressed, treat the illness with antidepressants (e. g. SSRIs), and withdraw the treatment when the patient has recovered–skills that were all used in this study. Finally, the investigators were their own controls and blind to the treatment. The bias that every rater has can be expected to be constant from one rating occasion to another, and, thus, cancelled out in the statistical analysis.

Reviewing the current evidence, one may have the impression that the results of this study disagree with the conclusions by Brenner and colleagues (2000) as well as Schrader (2000) and Harrer and colleagues (1999) who found that St. John's wort extract was as effective as SSRI drugs (sertraline and fluoxetine, respectively) in samples of mild to moderately depressed patients. None of these studies were placebo-controlled and, thus, were not able to demonstrate efficacy of any of these substances better than that of placebo. Our study also failed to show a difference in efficacy between *Hypericum* and fluoxetine. However, statistically significant superiority over placebo could not be demonstrated for any of the active substances, which is also in accordance with a recent study by the *Hypericum Depression Trial Study Group* (2002). To our knowledge, there are no large randomized and placebo-controlled studies of the efficacy of SSRI restricted to patients with mild to moderate depression.

Placebo effect

In recent years a number of clinical trials to test the efficacy of new potential antidepressants have failed to demonstrate superiority over placebo or the established reference drug. A rise in the response rates to placebo has not been paralleled by a rise in the response to the new substance (Montgomery 1999). One out of three trials fail to prove efficacy of a new antidepressant drug. Placebo response ranges between 25 and 55 %, whereas the effect of the investigational product and/or the active comparator ranges between 40 and 65 % in the same trials (The European Agency for the Evaluation of Medicinal Products 1998). In a meta-analysis, Walsh and colleagues have shown that the response to placebo in published trials of antidepressant medication for major depression is highly variable and often substantial and has increased significantly in recent years (Walsh et al. 2002).

The placebo effect is non-specific and can be effective in several ways, from spontaneous recovery to the increased attention the patient receives during a clinical trial. In a recent study, counseling seemed to be as effective as antidepressant treatment for mild to moderate depressive illness (Chilvers et al. 2001). There were no recommendations given to the investigators regarding how to keep the care constant from one center to another. A possible interpretation of the results is that the psychological care and the discrepancy in frequency of adverse events might explain the lack of difference between treatments. However, it is unlikely that the randomization procedure should favor the placebo group regarding psychological care in all 15 centers.

In this context, the long controversy regarding the efficacy of the SSRI must not be forgotten (Kirsch and Sapirstein 1998; Parker et al. 2003), and a possible publication bias of industry-sponsored studies (Baker et al. 2003).

Fixed doses

A fixed dosing regime may, however, have accounted for the failure of fluoxetine as well as *Hypericum*. A daily dose of 20 mg is recommended as a starting dose, but might not be optimal and might therefore be adjusted to a maximum daily dose of 80 mg (Stokes and Holtz 1997). However, in a study by Fabre and Putman (Fabre and Putman 1987), a mildly depressed group of patients showed no improvement at any dose level of fluoxetine (up to 60 mg). In a subgroup analysis of mildly depressed patients, Doogan and Langdon (1994) found no significant differences between sertraline and placebo.

Duration of treatment

It could be argued that the four weeks comparison period was too short. Relatively low response rates of about 40 % in all groups and remission rates around 25 % for both active treatments support this argument. Placebo response especially during the first 2–4 weeks of clinical trials may match active medication response due to the contribution of more unspecific factors of treatment response in the beginning, whereas separation of active drug from placebo may become more apparent after 6 and more weeks of treatment. Unfortunately, the ERC

did not approve a longer treatment with placebo than four weeks. However, also the reverse interpretation might be true, i. e. that the difference in remission rate between placebo and active treatment might disappear after the fourth week due to spontaneous recovery in the group of mildly depressed patients.

Interestingly, in the most recent placebo-controlled St. John's wort study by Shelton and colleagues (2001), which was not published at the time when the present study was conducted, the only significant difference between *Hypericum* and placebo regarding efficacy was found for remission rate (14 % vs. 5 %). The authors concluded that *Hypericum* was no more effective than placebo in their study. However, the lack of an active control group makes their conclusions premature. Despite treatment during eight weeks, only 19 % of placebo-treated patients responded to treatment. This unusually low percentage arouses suspicion that a substantial number of patients must have suffered from severe depression. Looking at the baseline demographics, this assumption was confirmed by the fact that 40 % of patients were diagnosed melancholic and that the mean duration of current major depressive disorder was 2.3 to 2.7 years. A more adequate conclusion from the results of the Shelton study might be that patients suffering from severe melancholic major depression probably cannot be helped with St. John's wort effectively, which is in accordance with the conclusions by Gupta and Möller (2003).

Compliance

The estimation of compliance in this study may be criticized. At every visit the number of pills and capsules was counted, which of course does not imply that the patient actually had taken the drug. It could be argued that this method should have been supplemented with a biomedical analysis of plasma levels. However, plasma samples are cross-sectional data reflecting compliance only for a short period depending on half-life of the substance assayed, and values usually show high interindividual variance. The lack of exact control of compliance is a common problem in most clinical trials.

Safety

The safety profile of St. John's wort extract was similar to that of placebo, and both treatments were significantly better tolerated than fluoxetine. Serious adverse events were not reported for any of the treatments. An underreporting of adverse events might have been the case, since no probe (according to e. g. the UKU, Lingjaerde et al. 1987) was conducted.

The favorable safety profile of *Hypericum* LI 160 is in accordance with the findings of other authors (Ernst et al. 1998; Linde et al. 2001; Stevinson and Ernst 1999). However, since none of the active treatments showed an efficacy better than that of placebo the higher tolerability of *Hypericum* LI 160 lacks clinical significance.

Acknowledgements The study was funded by a grant from Lichtwer Pharma GmbH, Berlin, Germany. We acknowledge the commitment of the 174 patients, the investigators and the staffs of the clinical study sites: Bratt I, Ehn L, von Zweigbergk D, Thingwall G (Solna); Hörnquist-Budell AM, Lindh A (Åkersberga); Åkerman L, Fagher B (Lund); Hanson M, Ödegaard G (Malmö); Hellner R, Movell B (Täby); Blomberg A, Lütz C (Gothenburg); Gabrielsson E (Stockholm); Babra M (Uppsala); Ljungström L (Simrishamn); Nielsen J (Skivarp); Nordlund T (Arlöv); Östgård P (Ystad).

References

1. American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). American Psychiatric Press: Washington DC
2. Baker CB, Johnsrud MT, Crismon ML, Rosenheck RA, Woods SW (2003) Quantitative analysis of sponsorship bias in economic studies of antidepressants. Br J Psychiatry 183:498–506
3. Brenner R, Azbel V, Madhusoodanan S, Pawlowska M (2000) Comparison of an extract of hypericum (LI 160) treatment of depression: a double-blind, randomized pilot study. Clin Ther 22:411–419
4. Brenner R, Bjerkenstedt L, Edman GV (2002) Hypericum perforatum extract (St. John's wort) for depression. Psychiatr Ann 32:21–26
5. Chilvers C, Dewey M, Fielding K, Gretton V, Miller PR, Palmer BW, Weller D, Churchill R, Williams I, Bedi N, Duggan C, Lee A, Harrison G (2001) Antidepressant drugs and generic counselling for treatment of major depression in primary care: randomised trial with patient preference arms. Br Med J 322:772–775
6. Deltito J, Beyer D (1998) The scientific, quasi-scientific and popular literature on the use of St. John's Wort in the treatment of depression. J Affective Disord 51:345–351
7. Doogan DP, Langdon CJ (1994) A double-blind, placebo-controlled comparison of sertraline and dothiepin in the treatment of major depression in general practice. Int Clin Psychopharmacol 9:95–100
8. Ernst E, Rand JI, Barnes J, Stevinson C (1998) Adverse effects profile of the herbal antidepressant St. Johns wort (*Hypericum perforatum* L.). Eur J Clin Pharmacol 54:589–594
9. The European Agency for the Evaluation of Medicinal Products (1998) Concept paper on the revision of the committee for proprietary medicinal products (CPMP) note for guidance on medicinal products for the treatment of depression (CPMP/EWP/518/97 concept paper). London: EMEA
10. Fabre LF, Putman HPr (1987) A fixed-dose clinical trial of fluoxetine in outpatients with major depression. J Clin Psychiatry 48:406–408
11. Gupta RK, Möller HJ (2003) St. John's Wort. An option for the primary care treatment of depressive patients? Eur Arch Psychiatry Clin Neurosci 253:140–148
12. Guy W (1976) ECDEU Assessment Manual for Psychopharmacology (76–338) Rockville, Md: US Department of Health, Education, and Welfare publication (ADM)
13. Hamilton M (1967) Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 6:278–296
14. Harrer G, Schmidt U, Kuhn U, Biller A (1999) Äquivalenzvergleich Johanniskrautextrakt LoHyp-57 versus Fluoxetin. Arzneim Forsch/Drug Res 49:3–10
15. Hoffmann J, Kühl ED (1979) Therapie von depressiven Zuständen mit Hypericin. Zeitschrift für Allgemeinmedizin 55:776–782
16. Hypericum Depression Trial Study Group (2002) Effect of Hypericum perforatum (St. John's Wort) in major depressive disorder: a randomized controlled trial. JAMA 287:1807–1814

17. Khan A, Leventhal RM, Khan SR, Brown WA (2002) Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. J Clin Psychopharmacol 22:40–45
18. Kim HL, Streltzer J, Goebert D (1999) St. John's wort for depression. A meta-analysis of well-defined clinical trials. J Nerv Ment Dis 187:532–539
19. Kirsch I, Sapirstein G (1998) Listening to Prozac but hearing placebo: a meta-analysis of antidepressant medication. Retrieved from the World Wide Web: http://journals.apa.org/prevention/volume1/pre0010002a.html
20. Linde K, Ramirez G, Mulrow CD, Pauls A, Weidenhammer W, Melchart D (1996) St John's wort for depression – an overview and meta-analysis of randomised clinical trials. Br Med J 313: 253–258
21. Linde K, ter Riet G, Hondras M, Vickers A, Saller R, Melchert D (2001) Systematic reviews of complementary therapies – an annotated bibliography. Part 2: Herbal medicine. BMC Complementary Alternative Medicine
22. Lingjaerde O, Ahlfors UG, Bech P, Dencker SJ, Elgen K (1987) The UKU side effect rating scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. Acta Psychiatr Scand 334:1–100
23. Montgomery SA (1999) The failure of placebo-controlled studies. Eur Neuropsychopharmacol 9:271–276
24. Montgomery SA, Åsberg M (1979) A new depression scale designed to be sensitive to change. Br J Psychiatry 133:382–389
25. Müller MJ, Himmerich H, Kienzle B, Szegedi A (2003) Differentiating moderate and severe depression using the Montgomery-Asberg depression rating scale (MADRS). J Affective Disord 77:255–260
26. Parker G, Anderson IM, Haddad P (2003) Clinical trials of antidepressant medications are producing meaningless results. Br J Psychiatry 183:102–104
27. Schrader E (2000) Equivalence of St John's wort extract (Ze 117) and fluoxetine: a randomized, controlled study in mild-moderate depression. Int Clin Psychopharmacol 15:61–68
28. Shelton RC, Keller MB, Gelenberg A, Dunner DL, Hirschfeld R, Thase ME, Russell J, Lydiard RB, Crits-Cristoph P, Gallop R, Todd L, Hellerstein D, Goodnick P, Keitner G, Stahl SM, Halbreich U (2001) Effectiveness of St John's wort in major depression: a randomized controlled trial. JAMA 285:1978–1986
29. Stevinson C, Ernst E (1999) Safety of Hypericum in patients with depression. CNS Drugs 11:125–132
30. Stokes PE, Holtz A (1997) Fluoxetine tenth anniversary update: the progress continues. Clin Ther 5:1135–1250
31. Walsh BT, Seidman SN, Sysko R, Gould M (2002) Placebo response in studies of major depression: variable, substantial, and growing. 287:1840–1847
32. World Health Organisation (1999) Adverse Reactions Terminology (WHO-ART). Uppsala: World Health Organisation
33. WMA (2000) Declaration of Helsinki. World Medical Association, www.wma.net/e/policty/17-c_e.html. Retrieved from the World Wide Web
34. Volz HP (1997) Controlled clinical trials of Hypericum extracts in depressed patients – an overview. Pharmacopsychiatry 30: 72–76
35. Volz HP, Laux P (2000) Potential treatment for subthreshold and mild depression: a comparison of St. John's wort extracts and fluoxetine. Compr Psychiatry 41:133–137
36. Vorbach EU, Arnoldt KH, Wolpert E (2000) St. Johns wort. A potential therapy for elderly depressed patients? Drugs Aging 16:189–197