Acute treatment of moderate to severe depression with hypericum extract WS 5570 (St John's wort): randomised controlled double blind non-inferiority trial versus paroxetine

A Szegedi, R Kohnen, A Dienel and M Kieser

BMJ 2005;330;503-; originally published online 11 Feb 2005; doi:10.1136/bmj.38356.655266.82

Updated information and services can be found at:
http://bmj.com/cgi/content/full/330/7490/503

These include:

References
This article cites 14 articles, 7 of which can be accessed free at:
http://bmj.com/cgi/content/full/330/7490/503#BIBL

8 online articles that cite this article can be accessed at:
http://bmj.com/cgi/content/full/330/7490/503#otherarticles

Rapid responses
27 rapid responses have been posted to this article, which you can access for free at:
http://bmj.com/cgi/content/full/330/7490/503#responses

You can respond to this article at:
http://bmj.com/cgi/eletter-submit/330/7490/503

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top left of the article

Correction
A correction has been published for this article. The contents of the correction have been appended to the original article in this reprint. The correction is available online at:
http://bmj.com/cgi/content/full/330/7494/759-a

Notes

To order reprints follow the “Request Permissions” link in the navigation box
To subscribe to BMJ go to:
http://resources.bmj.com/bmj/subscribers
Acute treatment of moderate to severe depression with hypericum extract WS 5570 (St John’s wort): randomised controlled double blind non-inferiority trial versus paroxetine

A Szegedi, R Kohnen, A Dienel, M Kieser

Abstract

Objective To investigate the efficacy of hypericum extract WS 5570 (St John’s wort) compared with paroxetine in patients with moderate to severe major depression.

Design Randomised double blind, double dummy, reference controlled, multicentre non-inferiority trial.

Setting 21 psychiatric primary care practices in Germany.

Participants 251 adult outpatients with acute major depression with total score ≥ 22 on the 17 item Hamilton depression scale.

Interventions 900 mg/day hypericum extract WS 5570 three times a day or 20 mg paroxetine once a day for six weeks. In initial non-responders doses were increased to 1800 mg/day hypericum or 40 mg/day paroxetine after two weeks.

Main outcomes Change in score on Hamilton depression scale from baseline to day 42 (primary outcome). Secondary measures were change in scores on Montgomery-Åsberg depression rating scale, clinical global impressions, and Beck depression inventory.

Results The Hamilton depression total score decreased by mean 14.4 (SD 8.9) points, corresponding to 56.6% (SD 34.3%) of the baseline value, in the hypericum group and by 11.4 (SD 8.8) points (44.8% (SD 33.5%) of baseline value) in the paroxetine group (intention to treat analysis; similar results were observed in the per protocol analysis). The intention to treat analysis (lower one sided 97.5% confidence limit 1.5 points for the difference hypericum minus paroxetine) and the per protocol analysis (lower confidence limit 0.7 points) showed non-inferiority of hypericum and statistical superiority over paroxetine. The lower limits in both cases exceeded the pre-specified non-inferiority margin of −2.5 points and the superiority margin of 0. The incidence of adverse events was 0.055 and 0.060 events per day of exposure for hypericum and paroxetine, respectively.

Conclusions In the treatment of moderate to severe major depression, hypericum extract WS 5570 is at least as effective as paroxetine and is better tolerated.

Introduction

Extract of Hypericum perforatum (St John’s wort) is more effective than placebo in the treatment of mild to moderate major depression and as effective as several tricyclic antidepressants or fluoxetine. In patients with more severe depression, however, the antidepressant efficacy of hypericum extract is disputed. In a comparison of 1800 mg/day hypericum extract (LI 160) and 150 mg/day imipramine the effect of both drugs was comparable during six weeks of acute treatment. That study, however, was not sufficiently powered to demonstrate non-inferiority of the herbal extract.

In clinical practice, hypericum extract is better tolerated than synthetic antidepressants. It may be particularly helpful in severe depression with its high risk of chronicity. We compared the efficacy and safety of hypericum extract with paroxetine in patients with moderate to severe depression.

Hypericum extract WS 5570 at a dose of 300 mg three times a day has been shown to be more effective than placebo in patients with mild to moderate major depression treated for six weeks. Paroxetine, on the other hand, is a potent selective serotonin reuptake inhibitor with proved efficacy in patients with depression of any severity and has a more favourable safety profile than tricyclic antidepressants. In major depression, daily doses between 20 mg and 50 mg have been recommended and are commonly used in clinical trials and in daily practice.

In accordance with Kaplan’s model of acute therapy and subsequent prophylactic treatment of unipolar depression, our study included a six week acute phase after which responders undergo four months of prophylactic continuation treatment (to prevent relapse or recurrence, or both).

Methods

Protocol, design, and objectives

This double blind, double dummy, randomised phase III trial examined the efficacy of hypericum extract WS 5570 compared with paroxetine in the acute treatment of moderate to severe major depression. After a screening examination participants underwent a single blind placebo run-in phase of three to seven days, during which they received three coated tablets of hypericum placebo per day plus one paroxetine placebo capsule in the morning. After that, we randomised those still meeting the selection criteria to six weeks of double blind treatment with hypericum extract or paroxetine. Those who responded to treatment (that is, their total score on the 17 item Hamilton depression scale decreased by ≥50%) were invited to participate in a four month double blind maintenance phase (reported elsewhere).

All patients provided written informed consent. We did not use a placebo control group because we considered it unethical to treat severely depressed patients with placebo for six weeks.

Participants

We recruited male and female outpatients in 21 psychiatric primary care centres in Germany. All participants were 18–70 years old and had single or recurrent moderate or severe episodes of
unipolar major depression without psychotic features (Diagnostic and Statistical Manual of Mental Disorders, fourth edition, (DSM-IV) 296.22, 296.23, 296.32, 296.33) persisting for two weeks to a year. At screening and baseline all patients had to have a total score ≥ 22 points on the 17 item Hamilton depression scale and ≥ 2 points for the item "depressive mood." The diagnosis of depression was based on the mini-international neuropsychiatric interview. There were no restrictions regarding ethnic group.

We excluded anyone with a decrease in total depression score of ≥ 25% during the run-in, or with a diagnosis of schizophrenia, acute anxiety disorder, adjustment disorder, depressive disorder of any type not stated above, bipolar disorder, organic mental disorder, acute post-traumatic stress disorder, or substance abuse disorder. We also excluded patients with increased risk of suicide (defined by a score ≥ 4 for item 10 of the Montgomery-Åsberg depression rating scale), who had previously attempted suicide, or who had not responded to more than one adequate treatment (equivalent to 150 mg/day amitriptyline for ≥ 6 weeks) in the present episode. Participants were not allowed to take other psychotropic medication and psychotherapy during the study (in case of previous antidepressant medication an appropriate wash out period of five half lives had to be observed).

Interventions and blinding
We used hypericum extract WS 5570 (Dr Willmar Schwabe Pharmaceuticals, Karlsruhe, Germany), a hydroalcoholic extract from herba hyperici (drug to extract ratio 3-7:1) with standardised contents of 3-6% hyperforin and 0.12-0.28% hypericin. The coated tablets contained 300 mg or 600 mg of the extract. Paroxetine was supplied in tablets of 20 mg packed in capsules containing one or two tablets. High and low dose tablets or capsules were indistinguishable in all aspects of their outward appearance. For each drug an identically matched placebo was dispensed to the centres in numbered containers. On inclusion of a patient into randomised treatment the local investigator allocated each participant the lowest available number. The block size was withheld from the investigators.

Random sequence generation, allocation concealment, implementation

Patients who still met the selection criteria at baseline were randomised at a ratio of 1:1 to hypericum or paroxetine. Randomisation was performed in blocks stratified by trial centre. A biometrician otherwise not involved in the trial generated the code using a validated computer program. The study drugs were dispensed to the centres in numbered containers. On inclusion of a patient into randomised treatment the local investigator allocated each participant the lowest available number. The block size was withheld from the investigators.

Statistical methods, sample size
Non-inferiority is usually established by showing that the true treatment difference is likely to be smaller than a prespecified non-inferiority margin that separates clinically important from clinically negligible (acceptable) differences. We considered that hypericum would not be relevantly inferior to paroxetine if the true decrease in total depression score (primary outcome measure) for hypericum was not more than 2.5 points smaller than for paroxetine (δ = −2.5).

The study was performed with an adaptive interim analysis. This design includes options for early stopping with rejection of the null hypothesis or for futility (boundaries $α_l = 0.01$ and $α_u = 0.5$, respectively) or for re-estimation of sample size in case of continuation.

For the change in total depression score we assessed non-inferiority of hypericum by a shifted $t$ test using the prespecified non-inferiority margin of 2.5 points and a global one sided type I error of $α = 0.025$. We used Fisher’s combination test in the final analysis, where the null hypothesis can be rejected when the product of the $P$ values from both study parts falls below $c_1 = 0.0038$. An analogous approach consists of calculating the one sided repeated 97.5% confidence limit for the treatment difference adjusted for the interim analysis. If this confidence limit is completely above the non-inferiority margin $δ = −2.5$, hypericum would be judged to be not inferior to paroxetine.

According to applicable guidance we reserved the option of testing for superiority after establishing non-inferiority of hypericum. If the lower one sided 95% confidence limit lies above 0, hypericum can be considered superior to paroxetine. We replaced missing values by carrying the last observation forward. The primary analysis was based on the intention to treat analysis to mirror clinical practice. We also performed a per protocol analysis to demonstrate robustness of the trial result to the choice of the analysis set. All secondary efficacy and safety measures were analysed descriptively. For the Hamilton total score, we defined response as a decrease in total score of ≥ 50% from baseline and remission as a score ≤ 10 points at week six.

We calculated the sample size for the first stage of the study until the interim analysis by assuming equal changes in depression score in each group with a common SD of 6 points. We needed $2 \times 50$ patients to attain 90% power for a one sided $P$ value of $P_c \leq 0.20$ in the interim analysis (trend towards non-inferiority of hypericum). The interim analysis resulted in a one sided $P_c = 0.084$ for the primary outcome measure so that the local type I error level for the second part of the trial was determined as $c_2 = 0.045$. Assuming a common SD of 6 points and equal means in both groups, we needed $2 \times 75$ patients to attain a power of 80% for the second stage of the trial, resulting in a total sample size of $2 \times 125$ patients.
Results

Participants

Between May 2000 and July 2003, we assessed 301 white patients and randomised and treated 251 (125 to hypericum and 126 to paroxetine). Figure 1 shows reasons for non-randomisation, premature termination, or exclusion. We did not exclude any patients because we thought they were at increased risk of suicide. Among the patients who were not randomised, two were withdrawn because they responded to placebo during the run-in period. All decisions regarding patient eligibility were made before code breaking.

Baseline demographic and clinical measures were comparable in both groups (table 1). Mean age and average duration of the current episode, however, were higher in the hypericum group. The baseline total depression scores ranged from 22 (minimum required) to 34 in both groups. In each group more than half of the patients had a total score \( \geq 25 \) and were thus severely depressed.¹⁴

Investigational treatment

After two weeks of randomised treatment, 69/122 patients in the hypericum group (57%) and 58/122 in the paroxetine group (48%) were switched to the higher doses. We assessed compliance with treatment by counting tablets; it was 96% (SD 7%) for hypericum and 98% (SD 10%) for paroxetine.

Figure 2 shows the total Hamilton depression scores over time. Between baseline and day 42 scores decreased by an average of 14.4 (SD 8.8) points (corresponding to 57% (SD 34%) of the baseline value) for hypericum and by 11.4 (SD 8.6) points (45% (SD 34%) for paroxetine (lower one sided repeated 97.5% confidence interval 16.5, 8.5).

| Table 1 Demographic and clinical characteristics at baseline (intention to treat analysis; figures are means (SD); medians unless stated otherwise) |
|-----------------------------------|-------------------|-------------------|
|                                   | Hypericum (n=122) | Paroxetine (n=122) |
| No (% of women)                   | 85 (70)           | 83 (68)           |
| Age (years)                       | 49.0 (11.0); 51.5 | 45.5 (11.5); 48.0 |
| No (% with recurrent depression)  | 50 (41)           | 49 (40)           |
| Duration of current episode (days) | 160 (109); 148    | 127 (81); 106     |
| HAMD total score*                 | 25.5 (2.7); 25.0  | 25.5 (2.9); 25.0  |
| No (% with HAMD total score \( \geq 25 \)) | 68 (53) | 67 (53) |
| MADRS total score†                | 29.9 (9.0); 29.0  | 29.4 (4.9); 29.0  |
| Beck depression inventory‡        | 26.3 (8.5); 26.0  | 25.6 (8.0); 24.5  |
| No (% markedly or severely ill§   | 87 (71)           | 84 (69)           |

HAMD=Hamilton depression scale. MADRS=Montgomery-Åsberg depression rating scale.

*Theoretical range 0–52.
†Theoretical range 0–60.
§According to clinical global impressions score.
confidence limit adjusted for the interim analysis\(^9\) for the difference hypericum–paroxetine was 1.5 points). In the per protocol analysis the decreases in scores during treatment were 14.6 (SD 9.0) points for hypericum and 12.0 (SD 8.5) points for paroxetine (lower confidence limit 0.7 points). Hence, the lower confidence limits not only exceeded the non-inferiority margin of −2.5 points but also the value 0, showing that hypericum is statistically superior to paroxetine at the one sided 2.5% level.

According to mean change in depression score from baseline, hypericum was descriptively superior to paroxetine in 11 of those 13 centres that had two or more patients in each group. At the end of the acute treatment phase 86/122 patients (71%) in the hypericum group and 73/122 (60%) in the paroxetine group responded to treatment (P = 0.08; \( \chi^2 \) test), and 61/122 (50%) and 43/122 patients (35%) showed remission (P = 0.02).

A subgroup analysis showed that patients who were switched from citalopram to hypericum or paroxetine because of lack of efficacy during the first two weeks of randomised treatment showed marked decreases in total depression score during weeks three to six. By the end of the double blind treatment period (day 42) we observed a substantial amelioration during weeks three to six. By the end of the double blind treatment period (day 42) we observed a substantial amelioration during weeks three to six. By the end of the double blind treatment period (day 42) we observed a substantial amelioration during weeks three to six.

Table 2 shows the main results for selected secondary measures. For all standardised psychiatric scales we found differences to day 42: hypericum 900 mg/day and 0.039 at 1800 mg/day) for hypericum and 0.060 (0.062 at 20 mg/day and 0.059 at 40 mg/day) for paroxetine (76%) reported 269 adverse events. The

### Table 2: Secondary measures (intention to treat analysis; figures are numbers (percentages) unless stated otherwise)

<table>
<thead>
<tr>
<th></th>
<th>Hypericum (n=122)</th>
<th>Paroxetine (n=122)</th>
<th>Difference (hypericum minus paroxetine) (95% CI), P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change from baseline to day 42:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADRS (mean (SD); median)†</td>
<td>16.4 (10.7); 17.0</td>
<td>12.6 (10.6); 14.0</td>
<td>3.8 (1.1 to 6.5), 0.09(^*)</td>
</tr>
<tr>
<td>BDI (mean (SD); median)†</td>
<td>10.2 (10.3); 9.0</td>
<td>7.0 (9.3); 5.5</td>
<td>3.2 (0.7 to 5.7), 0.01(^*)</td>
</tr>
<tr>
<td>Scores by day 42:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical global impressions:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1 much or very much improved</td>
<td>63 (66)</td>
<td>70 (57)</td>
<td>11 (-1 to 23), 0.09(^*)</td>
</tr>
<tr>
<td>Item 2 marked therapeutic effect</td>
<td>49 (40)</td>
<td>36 (30)</td>
<td>11 (-1 to 23), 0.08(^*)</td>
</tr>
<tr>
<td>Global efficacy self rating very good or good</td>
<td>65 (53)</td>
<td>55 (45)</td>
<td>8 (4 to 21), 0.20(^*)</td>
</tr>
</tbody>
</table>

### Discussion

**Principal findings**

We have shown that hypericum extract WS 5570 is at least as effective as paroxetine over six weeks of acute treatment in outpatients with moderate or severe unipolar major depression. This finding was stable across several validated investigator and self rating scales and across the participating centres as well as in different analysis datasets (including or excluding patients with major protocol violations). The average advantage of 3 points for the decrease in total Hamilton depression score from baseline underlines the clinical relevance of the observed effect,\(^9\) as do the responder rates of 70% vs 60% and the remission rates of 50% vs 35% for hypericum and paroxetine, respectively. The results thus indicate that in a group of patients in whom the appropriateness of hypericum extract was previously disputed, the antidepressant efficacy of the herbal drug is at least comparable with the effect of one of the leading synthetic antidepressants.

### Table 3: Adverse events that occurred in at least 10 patients in one group (safety analysis set; figures are numbers (percentages) of patients)

<table>
<thead>
<tr>
<th></th>
<th>Hypericum (n=125)</th>
<th>Paroxetine (n=126)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper abdominal pain</td>
<td>12 (9.6)</td>
<td>9 (7.1)</td>
</tr>
<tr>
<td>Diarrhoea</td>
<td>12 (9.6)</td>
<td>23 (18.3)</td>
</tr>
<tr>
<td>Dry mouth</td>
<td>16 (12.6)</td>
<td>25 (20.8)</td>
</tr>
<tr>
<td>Nausea</td>
<td>9 (7.2)</td>
<td>21 (16.7)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>14 (11.2)</td>
<td>16 (12.7)</td>
</tr>
<tr>
<td>Dizziness</td>
<td>9 (7.2)</td>
<td>24 (19.1)</td>
</tr>
<tr>
<td>Headache</td>
<td>13 (10.4)</td>
<td>14 (11.1)</td>
</tr>
<tr>
<td>Sleep disorder</td>
<td>5 (4.0)</td>
<td>10 (7.9)</td>
</tr>
<tr>
<td>Increased sweating</td>
<td>9 (7.2)</td>
<td>13 (10.3)</td>
</tr>
</tbody>
</table>
For paroxetine, 40 mg/day correspond to the established use of patients with insufficient response after two weeks of treatment. To our trial, which included a mandatory dose increase in paroxetine which was in line with previously published data from depression, especially as it is well tolerated.

Our results support the use of hypericum extract as an indicator of a pharmacological effect is that in both study groups could not be used in this group of predominantly severely depressed patients with insufficient response to the initial (lower) dose an (single blind) dose increase in initial non-responders was permitted. For the same reasons, we had to refrain from including patients at high risk of suicide. As we did not actually withdraw any patient because of increased risk of suicide, however, this restriction does not adversely affect the external validity of our data.

Implications for clinicians

Our results support the use of hypericum extract WS 5570 as an alternative to standard antidepressants in moderate to severe depression, especially as it is well tolerated. As in any effective antidepressant, potential interactions with other drugs deserve clinical attention.

The convincing results for hypericum extract WS 5570 observed in this trial deserve independent confirmation by other research. We are assessing efficacy in long term treatment, for which the drug can be an interesting option because of its favourable ratio of efficacy and tolerability, in the ongoing continuation phase.

We thank the investigators and patients, St Klein for project management, T Konstantinowicz for the data analysis, T Utz for project assistance, and A Volp for help with the manuscript.

Contributors: AS and RK conceived the study. AD conceived the study, and participated in its design and coordination. MK participated in the design of the study and was responsible for the analysis. All authors read and approved the final manuscript. AD and MK are guarantors.

Funding: Dr Willmar Schwabe Pharmaceuticals, manufacturer of WS 5570. Competing interests: AS has received consultancy fees from Dr Willmar Schwabe Pharmaceuticals. RK is head of a contract research organisation (IMEREM), which is engaged in several clinical trials of hypericum extract for different pharmaceutical companies. AD and MK are employees of Dr Willmar Schwabe Pharmaceuticals.

Ethical approval: The protocol was approved by the participating centres' appropriate independent ethics committees.

5. Wheeleys D. LI 168, an extract of St John's wort, versus amitriptyline in mildly to moderately depressed outpatients—a controlled f-week clinical trial. Pharmazie 1997;50(suppl 2):77-80.

(Accepted 17 December 2004)

doi 10.1136/bmj.38356.655266.82

Charité-Universitätsmedizin Berlin, Campus Benjamin Franklin, Department of Psychiatry and Psychotheraphy, Eschenallee 3, 14050 Berlin, Germany
A Szegedi managing senior psychologist
Institute for Medical Research Management and Biometrics GmbH, Scheuerstraße 21, 10047 Nürnberg, Germany
R Kohren head of scientific affairs
Dr Willmar Schwabe Pharmaceuticals, PO Box 410925, 76209 Karlsruhe, Germany
A Dienel head of clinical trials department
M Kieser head of biometry department
Correspondence to: M Kieser miki@schwabe.de
Commentary: Open access publishing: too much oxygen?

Jeffrey K Aronson

“We hold these truths to be self-evident . . .” This assertion of the US founding fathers betokened their zeal for human equality and rights. But such an attitude can betoken intellectual arrogance. It was, for example, self-evident to paediatricians in the 1950s that it would be beneficial to give premature babies 100% oxygen without proper trial. But 100% oxygen caused blindness, and the balance of benefit to harm was unfeasible.

In their survey of the attitudes of a small sample of scientists to open access1 Schroter and colleagues don’t actually trumpet its self-evident benefits, but their call for evidence refers to the author pays model, not open access publishing itself, although open access will not be possible without an author pays scheme or something comparable. But scientists’ opinions should not frame policy without supporting evidence. We need to ask whether immediate free access to readers, with whatever method of payment is used, would benefit science (not the scientists or the grant giving bodies) or society. To zealots (“the dream is now achievable”) the benefits of this 100% oxygen may be self-evident. But we have little evidence about the balance of benefits and harms. I believe that the potential advantages are few and the disadvantages many; I have summarised them on bmj.com.

Why should we uncritically adopt this system? We already have a better one, operated by many journals currently in and increasing numbers, in which readers pay for immediate access and access becomes universally free after a delay, for example 12 months, as required by the National Library of Medicine and the Wellcome Trust in their current initiative to digitise back issues of journals. Schemes such as HINARI (Health InterNetwork Access to Research Initiative) and AGORA (Access to Global Online Research in Agriculture) will maximise opportunities to access material that is published in this way.

In any system the burden of cost should be spread across those who are advantaged. A mixed model might be appropriate, maintaining subscriptions while allowing authors who want or are forced to pay for immediate free access to pay for it, and those who do not want it or cannot afford it, not to. Currently, some journals adopt author pays access, others do not. But there are many more readers than authors, which any balance in funding should reflect.

The uncritical application of basic values is a major source of unforeseen undesirable consequences of social actions.2 Who doesn’t instinctively feel that free access on day one is basically desirable? But we need to be completely sure that if we open the tap on the cylinder of this 100% oxygen the benefit to harm balance will be favourable, for we will not be able to turn the tap off—there will be no way back to subscription based journal publishing. As the third author of the above paper3 has written elsewhere, “think harm always.”4

Competing interests: JKA is a fellow of the British Pharmacological Society and chairman of the editorial board of the British Journal of Clinical Pharmacology, which is published by the society’s behalf by Blackwell Publishing, as a subscription journal with free access after 12 months: the complete archives of the journal are about to be digitised for free access.

2 Delamothe T, Smith R. Open access publishing takes off. The dream is now achievable. BMJ 2004;328:1-3.
5 Smith R. Think harm always [editor’s choice]. BMJ 2004;329. (3 July.)