# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEADER</td>
<td>1</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>1</td>
</tr>
<tr>
<td>PLAIN LANGUAGE SUMMARY</td>
<td>2</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>2</td>
</tr>
<tr>
<td>OBJECTIVES</td>
<td>3</td>
</tr>
<tr>
<td>METHODS</td>
<td>3</td>
</tr>
<tr>
<td>RESULTS</td>
<td>7</td>
</tr>
<tr>
<td>Figure 1</td>
<td>9</td>
</tr>
<tr>
<td>Figure 2</td>
<td>10</td>
</tr>
<tr>
<td>Figure 3</td>
<td>11</td>
</tr>
<tr>
<td>Figure 4</td>
<td>13</td>
</tr>
<tr>
<td>Figure 5</td>
<td>14</td>
</tr>
<tr>
<td>Figure 6</td>
<td>15</td>
</tr>
<tr>
<td>Figure 7</td>
<td>16</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>16</td>
</tr>
<tr>
<td>AUTHORS' CONCLUSIONS</td>
<td>18</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>19</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>19</td>
</tr>
<tr>
<td>CHARACTERISTICS OF STUDIES</td>
<td>25</td>
</tr>
<tr>
<td>DATA AND ANALYSES</td>
<td>49</td>
</tr>
<tr>
<td>WHAT'S NEW</td>
<td>53</td>
</tr>
<tr>
<td>HISTORY</td>
<td>54</td>
</tr>
<tr>
<td>CONTRIBUTIONS OF AUTHORS</td>
<td>54</td>
</tr>
<tr>
<td>DECLARATIONS OF INTEREST</td>
<td>55</td>
</tr>
<tr>
<td>SOURCES OF SUPPORT</td>
<td>55</td>
</tr>
<tr>
<td>INDEX TERMS</td>
<td>55</td>
</tr>
</tbody>
</table>
ABSTRACT

Background
In some countries extracts of the plant Hypericum perforatum L. (popularly called St. John's wort) are widely used for treating patients with depressive symptoms.

Objectives
To investigate whether extracts of hypericum are more effective than placebo and as effective as standard antidepressants in the treatment of major depression; and whether they have fewer adverse effects than standard antidepressant drugs.

Search strategy
Trials were searched in computerised databases, by checking bibliographies of relevant articles, and by contacting manufacturers and researchers.

Selection criteria
Trials were included if they: (1) were randomised and double-blind; (2) included patients with major depression; (3) compared extracts of St. John's wort with placebo or standard antidepressants; (4) included clinical outcomes assessing depressive symptoms.

Data collection and analysis
At least two independent reviewers extracted information from study reports. The main outcome measure for assessing effectiveness was the responder rate ratio (the relative risk of having a response to treatment). The main outcome measure for adverse effects was the number of patients dropping out due to adverse effects.

Main results
A total of 29 trials (5489 patients) including 18 comparisons with placebo and 17 comparisons with synthetic standard antidepressants met the inclusion criteria. Results of placebo-controlled trials showed marked heterogeneity. In nine larger trials the combined response rate ratio (RR) for hypericum extracts compared with placebo was 1.28 (95% confidence interval (CI), 1.10 to 1.49) and from nine smaller trials was 1.87 (95% CI, 1.22 to 2.87). Results of trials comparing hypericum extracts and standard antidepressants were statistically homogeneous. Compared with tri- or tetracyclic antidepressants and selective serotonin reuptake inhibitors (SSRIs), respectively, RRs were 1.02 (95% CI, 0.90 to 1.15; 5 trials) and 1.00 (95% CI, 0.90 to 1.11; 12 trials). Both in placebo-controlled trials and in comparisons with standard antidepressants, trials from German-speaking countries reported findings more favourable.
to hypericum. Patients given hypericum extracts dropped out of trials due to adverse effects less frequently than those given older antidepressants (odds ratio (OR) 0.24; 95% CI, 0.13 to 0.46) or SSRIs (OR 0.53, 95% CI, 0.34-0.83).

**Authors’ conclusions**

The available evidence suggests that the hypericum extracts tested in the included trials a) are superior to placebo in patients with major depression; b) are similarly effective as standard antidepressants; c) and have fewer side effects than standard antidepressants. The association of country of origin and precision with effects sizes complicates the interpretation.

**Plain Language Summary**

St. John's wort for treating depression.

Depression is characterised by depressed mood and/or loss of interest or pleasure in nearly all activities and a variety of other symptoms for periods longer than two weeks. Extracts of St. John's wort (botanical name *Hypericum perforatum* L.) are prescribed widely for the treatment of depression.

We have reviewed 29 studies in 5489 patients with depression that compared treatment with extracts of St. John's wort for 4 to 12 weeks with placebo treatment or standard antidepressants. The studies came from a variety of countries, tested several different St. John's wort extracts, and mostly included patients suffering from mild to moderately severe symptoms. Overall, the St. John's wort extracts tested in the trials were superior to placebo, similarly effective as standard antidepressants, and had fewer side effects than standard antidepressants. However, findings were more favourable to St. John's wort extracts in studies from German-speaking countries where these products have a long tradition and are often prescribed by physicians, while in studies from other countries St. John's wort extracts seemed less effective. This differences could be due to the inclusion of patients with slightly different types of depression, but it cannot be ruled out that some smaller studies from German-speaking countries were flawed and reported overoptimistic results.

Patients suffering from depressive symptoms who wish to use a St. John's wort product should consult a health professional. Using a St. John's wort extract might be justified, but important issues should be taken into account: St. John's wort products available on the market vary to a great extent. The results of this review apply only to the preparations tested in the studies included, and possibly to extracts with similar characteristics. Side effects of St. John's wort extracts are usually minor and uncommon. However, the effects of other drugs might be significantly compromised.

**Background**

**Description of the condition**

Depressive disorders are characterised by depressed mood and/or loss of interest or pleasure in nearly all activities in the presence of other symptoms such as loss of appetite, fatigue and lack of energy, sleep disturbance, restlessness or irritability, feelings of worthlessness or inappropriate guilt, difficulty in thinking, concentrating or making decisions and thoughts of death or suicide or attempts at suicide (Candy 2008). Depressive disorders are the largest source of non-fatal disease burden in the world, accounting for 12% of years lived with disability (Ustun 2004). There are two major classification systems to diagnose depressive disorders, the Diagnostic and Statistical Manual of Mental Disorders (DSM; current version DSM-IV-TR) and the International Statistical Classification of Diseases and Related Health Problems (ICD; current version ICD-10). DSM-IV defined depressive diagnoses include recurrent or persistent major depression and minor depression. ICD-10 diagnoses include recurrent or persistent depression with mild, moderate or severe episodes. According to the DSM-IV diagnostic classification, either depressed mood or a loss of interest or pleasure in daily activities consistently for at least a two week period has to be present to diagnose a major depressive disorder. The ICD-10 system uses the term depressive episode instead of major depressive disorder, but lists similar criteria.
**Description of the intervention**

Major depressive episodes are most commonly treated with antidepressant medication. Current first line drugs are selective serotonin reuptake inhibitors (SSRI) or tricyclic and related antidepressants (http://guidance.nice.org.uk/CG23). However, the size of effects over placebo in clinical trials has been modest (Turner 2008; Kirsch 2008), and although SSRIs are better tolerated than older antidepressants, side effects still occur in a relevant proportion of patients.

Extracts of the plant *Hypericum perforatum* L. (St. John’s wort), a member of the Hypericaceae family, have been used in folk medicine for a long time for a range of indications including depressive disorders. Extracts of St. John’s wort are licensed and widely used in Germany for the treatment of depressive, anxiety and sleep disorders. In recent years, hypericum extracts have also become increasingly popular in other countries.

**How the intervention might work**

The exact mechanism of action of the antidepressant effects of hypericum extracts is still unclear. Hypericum extracts contain at least seven constituents or groups of components that may contribute to its pharmacological effects (Nahrstedt 1997). These include naphthodianthrone (e.g., hypericins), flavonoids (e.g., quercetin), biflavonoids (e.g., biapigenin), xanthons, and phloroglucinol derivatives (e.g., hyperforin). Hypericum extracts have been shown to be active in a number of standard animal models that are used to indicate antidepressant effects (Wheatley 1998; Caccia 2005; Wurglics 2006). While some isolated substances, as for example hyperforin, have been shown to have antidepressant activity, the total extract seems to be more effective (Reichling 2003).

**Why it is important to do this review**

Hypericum extracts have been tested in a number of clinical trials since the 1980s. The first two versions of this review and other systematic reviews published between 1995 and 2000 concluded that hypericum extracts are more effective than placebo and are comparable to older antidepressants in the treatment of mild to moderate depressive disorders (Ernst 1995; Linde 1996; Linde 1998; Kim 1999; Gaster 2000; Williams 2000). Several trials included in these reviews were criticised because they included patients with few and/or mild symptoms who did not meet criteria for major depression, were conducted by primary care physicians who were not experienced in depression research, and/or used low doses of comparator drugs (Shelton 2001). In the 2005 update of our review (Linde 2005a; Linde 2005b) several new well-designed placebo-controlled trials were included, some of which had negative findings (Shelton 2001; HDTSG 2002) and which had spurred renewed debate about the efficacy of hypericum extracts.

We systematically investigated possible reasons for the contradictory findings. We found that larger, more precise studies yielded less positive results, suggesting that small studies with a higher risk of bias might overestimate the effects of hypericum extracts over placebo. The analyses also showed that effects over placebo were less pronounced in studies restricted to patients with major depression. Finally, we had the impression that studies originating from German-speaking countries (Germany, Austria, and Switzerland) had more positive results than studies originating from other countries independently from precision and formal diagnosis, although multiple regression analyses did not identify this as an independent predictor. Since we completed the search for our 2005 update, again, several new well-designed trials restricted to patients with major depression have been published. To sharpen the focus of this review, to reduce clinical heterogeneity, and to reflect the fact that almost all new high-quality trials of hypericum extracts are restricted to patients with major depression, we decided to limit the review now to this group of patients.

**OBJECTIVES**

This updated review aimed to investigate whether extracts of hypericum:

- are more effective than placebo and
- as effective as standard antidepressant drugs, and
- whether they have less adverse effects compared to standard antidepressant drugs

in the treatment of major depression in adults.

In addition, we investigated possible reasons for varying results across studies, with a focus on precision of the studies, baseline severity of depression, and country of origin.

**METHODS**

Criteria for considering studies for this review

Types of studies

To be included trials had to be double-blind and randomised.
Types of participants
Patients had to suffer from major depression (meeting DSM-IV or ICD-10 criteria). Trials in children (< 16 years) were not eligible. In previous versions of this review (Linde 1998; Linde 2005a) trials not restricted to patients with major depression had been included.

Types of interventions
Experimental intervention
Preparations of hypericum (St. John’s wort). Trials investigating combinations of hypericum with other herbs were excluded.

Control intervention
Placebo or synthetic antidepressants (tricyclic and related antidepressants, selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors). Trials using clearly inadequate synthetic antidepressants (e.g., benzodiazepines) or a dosage clearly below the lower thresholds recommended in current guidelines (Härter 2003, ICSI 2007) were excluded.

Experimental and control treatments had to be given for at least four weeks.

Main comparisons
The following comparisons were performed:
1. hypericum extracts vs. placebo
2. hypericum extracts vs. standard antidepressants

Types of outcome measures
Primary outcome
To be included, trials had to measure clinical outcomes such as depression scales or symptoms. Trials that measured physiological parameters only were excluded. The primary outcomes of interest were
1. Effectiveness: treatment response
2. Safety: the proportion of patients who dropped out due to adverse effects

Secondary outcomes
1. Effectiveness: remission, depression scales such as the Hamilton Depression Scale (HAMD), the Clinical Global Impression Index (CGI), the Montgomery-Asberg Depression Rating Scale (MADRS), patient-rated depression scales
2. Safety: total proportion of drop-outs, proportion of patients reporting adverse effects

Search methods for identification of studies
For the first version of the review we searched for published and unpublished eligible trials in the following ways:
1. Electronic searches
   a) Clinical Trials Register of the Cochrane Collaboration Depression, Anxiety & Neurosis Group (CCDANTR)
   b) database of the Cochrane Field for Complementary Medicine
   c) full text searches in MedlineSilverPlatter CD-ROM from 1983 onwards and Embase 1989 onwards using the terms ‘St. John’s wort’, ‘Johanniskraut’ (German for St John’s wort), ‘hyperic*’
   d) full text searches in Psychlit and Psychindex 1987 - 1997 CD-ROM
   e) searches in the private database Phytodok, Munich.
2. Searching other resources
   a) Checking bibliographies of obtained articles
   b) Contacting pharmaceutical companies and authors.

There were no language restrictions.

For the updated version of the review, we searched for published and unpublished eligible trials in the following ways:
1. Electronic searches
   For the update, regular electronic searches were performed in CCDANTR (last search July 2007) and PubMed (screening all hits for text word “hypericum”, last search July 8, 2008).
2. Searching other resources
   We screened bibliographies of published articles, and repeatedly contacted experts, researchers, and manufacturers inquiring for new trials. One reviewer (KL) initially screened reference lists to identify controlled studies on hypericum preparations in humans. All possibly relevant studies or publications were then checked formally for eligibility.

Data collection and analysis
Selection of studies
Two reviewers independently decided on eligibility for the revised inclusion criteria. Disagreements were resolved through discussion. Due to reading errors, disagreements occurred for two trials in which not all patients had major depression and which were excluded after discussion (Vorbach 1994, Winkel 2000). In two trials both reviewers had problems with assessing eligibility: For one small, older trial (Lehrl 1993) the publication did not state that inclusion was limited to patients with major depression, but we had a statement of the sponsor obtained for our 1998 update that all patients met the criteria. As this information could not be verified for this update, we decided to exclude the trial. A Chinese trial (Gu 2001) referred to a Chinese classification. As this classification is not completely comparable to ICD-10 and DSM-IV, we decided to also exclude this trial.

Data extraction and management
Primary study characteristics and results were extracted by at least two independent reviewers using a pretested form. In particular, we extracted diagnoses and main inclusion criteria, age, gender, duration of episodes, baseline depression scores, country of origin, number and type of study centers, numbers of patients who were randomised and analysed and who completed protocols, the
number and reasons for drop-outs and withdrawals, numbers of patients reporting adverse effects, and the number and type of adverse effects that were reported. We assessed numbers of patients who were classified as responders based on score improvements on the Hamilton Rating Scale for Depression (HAMD), the Clinical Global Impression Index (CGI; subscale global improvement rating as at least “much improved”), or any other clinical response measurement. Missing or additional information was sought from authors/sponsors. Most trials measured clinical outcomes with the Hamilton Depression Scale (HAMD) and the Clinical Global Impression Index (CGI). The HAMD is an observer-rated scale that focuses mainly on somatic symptoms of depression (Hamilton 1960). The original version includes 21 items, but a version with 17 items is more commonly used in clinical trials. Most studies using the HAMD report the number of ‘treatment responders’ (patients achieving a score less than 10 and/or less than 50% of the baseline score). When available, we extracted means and standard deviations before, during and after treatment as well as the number of ‘responders’. The CGI (CGI 1970) is an observer rated instrument with three items (severity of illness, global improvement, and an efficacy index). We extracted the number of patients rated as ‘much improved’ or ‘very much improved’ for global improvement. As recently the Montgomery-Asberg Depression Rating Scale (MADRS (Montgomery 1979)) and the remission criterion for the HAMD (usually a score of less than 8 at the end of treatment) have been gaining importance as outcome criteria, we also checked all trials for the reporting of these measures. As the DS (Depression Scale von Zerssen (von Zerssen 1996)) was the most often used patient-rated instrument in the included trials, we extracted post-treatment data for this scale, if available. For additional post-hoc analyses, one reviewer (KL) also extracted data for other self-rating instruments.

Assessment of risk of bias in included studies
The main part of the update process of this review was completed before the new risk of bias tool of the Cochrane Collaboration (Higgins 2008) was available. The methodological quality of each trial was assessed by at least two independent reviewers using scales developed by A. Jadad et al. (Jadad 1996) and by one of the reviewers (KL). The results of the quality scoring are displayed in the table of included studies. The Jadad scale has three items adding up to a maximum score of five points. 0, 1 or 2 points can be given for randomisation (explicit statement that allocation was randomised and description of an adequate generation of the random sequence), 0, 1 or 2 points for double-blinding (explicit statement that patients and evaluators were blinded and that treatments were indistinguishable), 0 or 1 point for description of drop-outs and withdrawals (numbers and reasons for all compared groups separately). The display in the table of included studies is as follows (examples): 2-2-1 (full score in every item), 1-0-0 (only statement on randomisation). The second quality scale, the “Internal Validity Scale” (IV), which has been used in other reviews on complementary medicine (Linde 1996b, Linde 1997) has six items with possible scores of 0, 0.5 or 1 point for each. Items 1 through 6 refer to statement of random allocation, adequacy of randomization concealment, baseline comparability, blinding of patients, blinding of evaluators, and likelihood of selection bias after allocation, respectively. Results are displayed by item in the table of included studies (e.g., 1-1-1-0.5-1-1 represents a full score, with the exception of blinding of patients which was stated but treatment and placebo might have been distinguishable).

The assessments in the Jadad and IV scores are solely based on the information provided in the publication (as additional information could not be gathered for all studies). In the table 'Characteristics of included studies', however, additional information provided from authors or sponsors was included. This table also contains information on allocation concealment and attrition.

Measures of treatment effect
Our primary outcome measure, to assess the effectiveness of St John’s wort versus placebo and versus other antidepressants, was the proportion of responders (according to the Hamilton Depression Scale (first preference) or other responder measures (second preference)) at the end of treatment, or in case of treatment phases longer than 6 weeks, at the time point defined for primary outcome measurement by the study investigators. Secondary outcome measures were: proportion of responders according to HAMD, proportion of responders according to CGI, mean HAMD after treatment (or, if this was not available, difference after treatment - baseline), at 2, 4, 6 to 8 weeks, and mean DS score after treatment (or, if this was not available, difference after treatment - baseline).

The main outcome measure for the safety analysis was the proportion of patients who dropped out due to adverse effects. Secondary measures were the total proportion of drop-outs and the proportion of patients reporting adverse effects.

Dichotomous outcomes
We used responder rate ratios (= relative risks = proportion of responders in the treatment group/proportion of responders in the control group) and their 95% confidence intervals for the analysis of treatment response. Responder rate ratios greater than 1 indicate better response in the hypericum group. Due to highly variable frequency of side or adverse effects reported, odds ratios instead of rate ratios were calculated in the safety analyses. Odds ratios less than 1 indicate that fewer events occurred in the hypericum group. Continuous outcomes
For HAMD and DS scores we calculated mean differences (also termed weighted mean differences). Negative mean differences indicate better response in the hypericum group.
Unit of analysis issues
Two trials with more than one hypericum group were included in the analyses as follows: Laakmann 1998 included an extract available on the market and an additional experimental extract with low hyperforin content which was never on the market and only used for control reasons. We did not include the data from the group receiving the experimental low-hyperforin extract in the analyses. Kasper 2006 et al tested two dosages (600 and 1200 mg) of an available product. We pooled the data from these two groups to prevent that the control group of this trial would have been included in the analyses twice.

Dealing with missing data
Dichotomous outcomes
Responder proportions were calculated according to the intention to treat principle, counting drop-outs as non-responders. For the comparison hypericum extracts vs. standard antidepressants responder proportions were also calculated on a per protocol basis (as this is considered more appropriate to assess the equivalence of two treatments).

Continuous outcomes
If means and standard deviations from intent to treat analysis with missing values replaced were available, we preferably used these data. In other cases we used analysis based on available data.

Obtaining missing data
If the number of patients responding to treatment and means and/or standard deviations of HAMD scores after completion were not reported, we always tried to contact first or corresponding authors and/or sponsors to obtain these data. In general, we also tried to obtain other missing details on methods and secondary outcomes from authors or sponsors, but the extent to which we were doing this depended on the cooperation of authors/sponsors and the amount of missing information in the publications. We did not impute or recalculate missing standard deviations as these were unavailable only for a few secondary outcomes in a minority of trials. We tried to contact authors and/or sponsors of 27 of the 29 included trials; for two trials (Kalb 2001; Laakmann 1998) this was considered unnecessary. We did not receive responses for five trials (Behnke 2002; Brenner 2000; Fava 2005; Harrer 1999; Moreno 2005). Very limited additional information was available or needed for three studies (Bjerkenstedt 2005; Volz 2000; Woelk 2000). We obtained relevant additional information to a variable extent from authors, sponsors, or both for the remaining 19 trials.

Data synthesis
The following comparisons were performed:
1. hypericum extracts vs. placebo: a) for dichotomous outcomes (response rate ratios); b) for continuous outcomes; c) for drop-outs and adverse effects
2. hypericum extracts vs. standard antidepressants: a) for dichotomous outcomes; b) for continuous outcomes; c) for drop-outs and adverse effects

All main analyses were performed using RevMan 5. Due to the clinical diversity of the studied populations, the hypericum extracts and the comparison drugs used, we considered that the included studies did not estimate a common underlying effect, but rather that each individual study estimated its single and unique underlying effect. Thus, the application of random effects model in all analyses seemed to be appropriate.

The primary analysis for the comparison of response rate ratios (= relative risks) under treatment with hypericum extracts or placebo was a random-effects intention to treat meta-analysis stratified by study precision (above or below median of variance of treatment effect).

Assessment of heterogeneity
Heterogeneity of trials’ results was tested with the Chi-squared test, and the I-squared statistic was calculated to give an estimate of the degree of heterogeneity. I-squared values over 50% indicate considerable heterogeneity (Higgins 2003).

Investigation of heterogeneity and subgroup analyses
Predefined subgroup analyses were performed (a) including only trials with response operationalised with the HAMD score; (b) including only trials with response operationalized with the CGI; (c) for the type of extract investigated; and (d) comparing trials originating from German-speaking countries and from other countries. Weighted mean differences for HAMD scores were calculated after therapy, at 2 to 3, 4, 6 to 8 weeks, and for differences compared to baseline values. For DS scores we calculated after therapy values, for MADRS score after therapy values and differences compared to baseline. As only relatively few studies used the DS, we performed an additional post-hoc random effects analysis calculating standardised mean differences for any available patient-rating scale (preferably end of treatment values, but if these were not available also differences from baseline) to investigate whether findings from physician-rated instruments could be broadly reproduced. The primary analysis for the comparison of responder rate ratios under treatment with hypericum extracts or standard antidepressants was a random effects intent to treat meta-analysis stratified for type of synthetic antidepressant (selective serotonin inhibitors or other antidepressants). Predefined subgroup analyses were performed (a) using per protocol data; (b) stratified for country (German-speaking Europe versus other countries); (c) including only trials with response operationalised with the HAMD score; and (d) including only trials with response operationalised with the CGI.

Additional meta-regression analyses were performed to investigate the influence of country of origin (German-speaking versus not German speaking), precision and HAMD baseline values on study findings (responder ratio and mean difference of HAMD scores after treatment) both in placebo and standard antidepressants comparisons. According to current recommendations of ex-
The results of this meta-analysis are as follows:

**Description of studies**

See: Characteristics of included studies; Characteristics of excluded studies.

**Results of the search**

A total of 79 possibly relevant studies were identified and checked formally for eligibility.

**Included studies**

Twenty nine trials including a total of 5489 (range 30 to 388) patients met inclusion criteria (see Characteristics of included studies).


**Types of intervention**

A variety of hypericum preparations were studied in the trials. The range of daily extract doses varied between 240 and 1800 mg, but in most trials 500 to 1200 mg were used. The standard antidepressants used as active comparators were fluoxetine (6 trials, dosage 20 to 40 mg), sertraline (4 trials, 50 to 100 mg), imipramine (in 3 trials, dosage 100 to 150 mg), citalopram (1 trial, 20 mg), paroxetine (1 trial, 20 to 40 mg), mirtazapine (1 trial, 15 mg), and amitriptyline (1 trial, 75 mg). The comparator dosage of mirtazapine and amitriptyline were slightly below of those recommended in current guidelines (Härter 2003, ICSI 2007) and in most other studies at the minimum of recommended dosages. The treatment periods lasted 4 (1 trial), 6 (19), 7 (1), 8 (5) or 12 weeks (4 trials). Four trials included some long-term follow-up or continuation treatment after the main trial phase (Brenner 2000; Gastpar 2005; Shelton 2001; Szegedi 2005).

**Types of outcome**

The most frequently used instrument used for outcome measurement was the Hamilton Rating Scale for Depression (used in all trials). A variety of other ratings scales and instruments were used in addition.

**Excluded studies**

Fifty trials (see Characteristics of excluded studies) did not meet the criteria for inclusion in this meta-analysis.
inclusion criteria: eight trials were not limited to patients with depression (Albertini 1986; Bendre 1980; Dittmer 1992; Hottenrott 1997; Maisenbacher 1995; Panijel 1985; Sindrup 2000; Volz 2002), four trials were on prevention or treatment of depressive symptoms in patients suffering primarily from other diseases (Haring 1996; Li 2005; Mo 2004; Werth 1989), two measured physiological parameters only (as EEG) in depressed patients (Czekalla 1997; Kugler 1990b), five did not include a placebo or standard drug comparison group (Bernhadt 1993; Lenoir 1999; Martinez 1993; Spielberger 1985; Zeller 2000), eight involved healthy volunteers (Brockmöller 1997; Herberg 1992; Johnson 1992; Johnson 1993; Schmidt 1993b; Schulz 1993; Staffeldt 1993; Wienert 1991), three tested combinations of hypericum and other herbal extracts (Ditzler 1992; Kniebel 1988; Steger 1985), and two compared hypericum extract with medications which are no longer considered adequate for depression (diazepam or bromazepam) (Kugler 1990a; Warnecke 1986); one of these trials also was not explicitly randomized. Due to the new exclusion criterion, we excluded 17 trials not restricted to patients with major depression. Fifteen had been included in the previous version of the review (Halama 1991; Harrer 1991; Hoffmann 1979; Hübner 1993; König 1993; Lehrl 1993; Osterheider 1992; Quandt 1993; Reh 1992; Schlich 1987; Schmidt 1989; Schmidt 1993; Sommer 1994; Vorbach 1994; Winkel 2000) while two were not (Agrawal 1994, for which it had not been possible to obtain a full copy and Gu 2001, which was newly identified in the update searches). Finally, we excluded one previously included trial as the standard antidepressant treatment was far below recommended dosages (30 mg Amitriptyline daily; Bergmann 1993).

Risk of bias in included studies
The majority of the trials were of high quality. The median quality scores were 5 (out of 5, range 2 to 5) for the Jadad scale and 4.5 (out of 6; range 2 to 6) for the IV scale (see quality rating of the single trials in the Characteristics of included studies).

Sequence generation/allocation concealment
The information on how the random sequence was generated was reported or provided on request for 18 trials (in all cases a computer program). Twenty two trials reported an adequate method of allocation concealment (most often consecutively numbered medication).

Blinding
All trials were described as double-blind, but only one trial reported that blinding was tested (HDTSG 2002). In this three-armed trial (hypericum vs. sertraline vs. placebo) about a third (as expected by chance alone) of guesses made by physicians were correct for hypericum and placebo patients, but in 66% of sertraline patients (p = 0.001).

Incomplete outcome data
In some trials attrition rates were high (for example, Fava 2005; see Characteristics of included studies). All placebo-controlled trials included an intent to treat analysis.

Effects of interventions

Comparison I: Hypericum extracts versus placebo

1. Effectiveness
a) Responder analyses
Sixteen of the 18 placebo-controlled trials reported the number of patients classified as responders based on score reduction on the HAMD scale, one trial reported response according to the MADRS scale (Bracher 2001), and one trial only reported the proportion of patients rated at least as “improved” for the CGI (Volz 2002). Patients receiving hypericum extracts were significantly more likely to be responders (RR = 1.48; 95%CI 1.23 to 1.77; see comparison 1.1 and Figure 1) but study results were highly heterogeneous ($I^2 = 75$%). Effects in favour of hypericum extracts were less pronounced in more precise trials (RR = 1.28; 95%CI 1.10 to 1.49) compared to less precise trials (RR = 1.87; 95%CI 1.22 to 2.87) but heterogeneity was still strong in both subgroups ($I^2 = 61$% and 79%, respectively).
Findings were similar if response rates were based only on the trials reporting response according to the HAMD scale, or on the CGI (see comparisons 1.2 and 1.3). If trials investigating defined extracts were analysed separately (subgroups of trials testing the same extracts; see comparison 1.4), heterogeneity was strong in 3 of 4 subgroups.

Trials from German-speaking countries reported more positive findings than trials from other countries (RR = 1.78; 95% CI 1.42 to 2.25 vs. 1.07; 95% CI 0.88 to 1.31, respectively; see comparison 1.5 and Figure 2). Six trials reported remission rates. These were significantly higher in patients receiving hypericum extracts than in those receiving placebo (RR = 2.77; 95% CI 1.80 to 4.26; I² = 29%; see comparison 1.6).
Figure 2. Forest plot of comparison: 1 Hypericum mono-preparations vs. placebo A. Dichotomous measures, outcome: 1.5 Responder among studies from German-speaking countries and other studies.

There was significant funnel plot asymmetry for the main responder analysis (coefficient = 2.19, p = 0.03; Figure 3).
In univariable meta-regression analyses, country of origin (studies from German-speaking countries showing larger effect sizes; \( p = 0.002 \)), precision (more precise studies showing smaller effects; \( p = 0.032 \)) and baseline values (higher values associated with smaller effect sizes; \( p = 0.048 \)) were significantly associated with effects sizes. In multiple analyses the association remained significant for country of origin (\( p = 0.035 \)) and precision (\( p = 0.017 \)) but became non-significant for HAMD baseline values. Altogether over half of the variance (\( R^2 = 0.51 \)) could be explained by these three variables. Findings of meta-regression analyses are summarised in Appendix 1.

**b) Analyses of depression scales**

Analyses based on mean HAMD values yielded similar findings. At the completion of treatment HAMD values were 3.04 (95%CI 1.78 to 4.29) score points lower in hypericum groups compared to placebo groups, but there was strong heterogeneity (\( I^2 = 86\% \); see comparison 2.1). Effects over placebo were significant after 2 (comparison 2.2), 4 (comparison 2.3), 6 to 8 weeks of treatment (comparison 2.4), and for changes from baseline to end of treatment (comparison 2.5). Significant effects over placebo were also reported for the MADRS (comparisons 2.6, and 2.7). Studies from German-language countries reported much larger effects over placebo (weighted mean difference = 4.29, 95%CI 2.97 to 5.61 score points; comparison 2.8) than studies from other countries (MD = 0.77 score points, 95%CI 0.20 to 1.74 score points). There was no significant funnel plot asymmetry (coefficient in the analysis of HAMD values at completion of treatment = -2.12, \( p = 0.35 \)).

In multiple meta-regression analysis, country of origin was significantly associated with effects size (larger effects in trials from German-speaking countries; \( p < 0.001 \)) but not precision and HAMD baseline values (\( R^2 = 0.63 \); see Appendix 1).

The four trials reporting results for the patient-rated von Zerssen Depression Scale (D-S) showed a significant effect of hypericum extracts over placebo (comparison 2.9). Post-hoc analyses using available data from 12 placebo-controlled trials for a variety of self-rating instruments also confirmed analyses based on physician-rated outcomes. The pooled standardised mean difference (SMD) was -0.47 (95% CI -0.64 to -0.30; \( I^2 = 74\% \); see comparison 2.10). Trials from German-speaking countries again reported more favourable findings than trials from other countries (SMDs of -
0.57 and -0.17 respectively; see comparison 2.11).

2. Safety

Primary outcome
The number of patients dropping out for adverse effects was similar among patients receiving hypericum extracts and placebo (OR = 0.92, 95%CI 0.45 to 1.88, I²=0%; see comparison 3.1).

Secondary outcomes
The total number of patients dropping out and the number of patients dropping out for any reason were similar among patients receiving hypericum extracts and placebo (comparisons 3.2 and 3.3).

Comparison 2: Hypericum extracts versus standard antidepressants

1. Effectiveness

a) Responder analyses
All 17 trials comparing hypericum extracts to standard antidepressant treatment reported the number of responders according to the HAMD score. Based on an intention to treat approach the pooled responder rate ratio was 1.01 for all 17 trials (95%CI 0.93 to 1.09; I² = 17%; see comparison 4.1 and Figure 4). For the five trials comparing hypericum extracts with older antidepressants, the pooled estimate was 1.02 (95%CI 0.90 to 1.15; I² = 0%), and 1.00 for the 12 trials with selective serotonin reuptake inhibitors (95%CI 0.90 to 1.12; I² = 29%).
Figure 4. Forest plot of comparison: 4 Hypericum mono-preparations vs. standard antidepressants. A. 
Dichotomous measures, outcome: 4.1 Responder (intent to treat) - primary analysis.

If analyses were based on per protocol data, the pooled responder rate ratio was 0.96 (95% CI 0.88 to 1.05; I² = 43%; see comparison 4.2). Analysis based on the CGI also found no relevant differences (RR = 1.04; 95% CI 0.94 to 1.09; I² = 24%; see comparison 4.3). In trials originating from German-speaking countries findings were slightly more favourable to hypericum than in trials from other countries (RR 1.04 and 0.90, respectively; see comparison 4.4 and Figure 5). In the four trials reporting remission rates the response rate ratio was 1.24 (95% CI 1.02 to 1.50; I² = 0%; see comparison 4.5).
Figure 5. Forest plot of comparison: 4 Hypericum mono-preparations vs. standard antidepressants. A. Dichotomous measures, outcome: 4.4 Responder among studies from German-speaking studies and other studies.

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Hypericum Events</th>
<th>Control Events</th>
<th>Total Events</th>
<th>Total Weight</th>
<th>M.H. Random, 95% CI</th>
<th>Risk Ratio M.H. Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastpar 2005</td>
<td>70</td>
<td>102</td>
<td>72</td>
<td>56</td>
<td>12.8%</td>
<td>0.93 [0.78, 1.12]</td>
</tr>
<tr>
<td>Gastpar 2006</td>
<td>71</td>
<td>131</td>
<td>71</td>
<td>127</td>
<td>9.3%</td>
<td>0.97 [0.78, 1.21]</td>
</tr>
<tr>
<td>Hamer 1993</td>
<td>27</td>
<td>51</td>
<td>28</td>
<td>51</td>
<td>4.3%</td>
<td>0.96 [0.77, 1.20]</td>
</tr>
<tr>
<td>Hamer 1999</td>
<td>50</td>
<td>77</td>
<td>57</td>
<td>84</td>
<td>0.3%</td>
<td>0.96 [0.77, 1.19]</td>
</tr>
<tr>
<td>Phillips 1998</td>
<td>76</td>
<td>105</td>
<td>78</td>
<td>110</td>
<td>11.4%</td>
<td>1.13 [0.94, 1.36]</td>
</tr>
<tr>
<td>Schradar 2000</td>
<td>57</td>
<td>125</td>
<td>39</td>
<td>113</td>
<td>5.3%</td>
<td>1.32 [0.98, 1.82]</td>
</tr>
<tr>
<td>Szegedi 2005</td>
<td>86</td>
<td>122</td>
<td>83</td>
<td>122</td>
<td>11.9%</td>
<td>1.18 [0.98, 1.42]</td>
</tr>
<tr>
<td>Voller 1997</td>
<td>38</td>
<td>107</td>
<td>41</td>
<td>102</td>
<td>4.3%</td>
<td>0.84 [0.69, 1.02]</td>
</tr>
<tr>
<td>Vibol 2000</td>
<td>88</td>
<td>157</td>
<td>57</td>
<td>157</td>
<td>7.4%</td>
<td>1.00 [0.83, 1.40]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>978</td>
<td>974</td>
<td>76.0%</td>
<td>1.04</td>
<td>[0.96, 1.13]</td>
<td></td>
</tr>
</tbody>
</table>

Total events 541, 518
Heterogeneity Tau² = 0.00; Chi² = 0.67; df = 8 (P = 0.97); I² = 9%
Test for overall effect Z = 0.66 (P = 0.24)

4.4.2 Studies from other countries

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Hypericum Events</th>
<th>Control Events</th>
<th>Total Events</th>
<th>Total Weight</th>
<th>M.H. Random, 95% CI</th>
<th>Risk Ratio M.H. Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eibensteiner 2002</td>
<td>16</td>
<td>35</td>
<td>21</td>
<td>35</td>
<td>2.8%</td>
<td>0.76 [0.49, 1.16]</td>
</tr>
<tr>
<td>Bjerkenstig 2005</td>
<td>22</td>
<td>54</td>
<td>20</td>
<td>54</td>
<td>2.8%</td>
<td>1.10 [0.88, 1.33]</td>
</tr>
<tr>
<td>Brenner 2000</td>
<td>7</td>
<td>15</td>
<td>8</td>
<td>15</td>
<td>0.9%</td>
<td>1.17 [0.51, 2.46]</td>
</tr>
<tr>
<td>Fava 2005</td>
<td>17</td>
<td>45</td>
<td>14</td>
<td>47</td>
<td>1.1%</td>
<td>1.27 [0.71, 2.32]</td>
</tr>
<tr>
<td>HOLL 2002</td>
<td>48</td>
<td>113</td>
<td>55</td>
<td>108</td>
<td>8.1%</td>
<td>0.91 [0.62, 1.33]</td>
</tr>
<tr>
<td>Merz 2005</td>
<td>4</td>
<td>20</td>
<td>11</td>
<td>20</td>
<td>0.7%</td>
<td>0.36 [0.14, 0.95]</td>
</tr>
<tr>
<td>van Gorp 2002</td>
<td>25</td>
<td>45</td>
<td>23</td>
<td>45</td>
<td>3.7%</td>
<td>1.09 [0.74, 1.60]</td>
</tr>
<tr>
<td>Vollet 1997</td>
<td>40</td>
<td>87</td>
<td>42</td>
<td>78</td>
<td>5.5%</td>
<td>0.85 [0.63, 1.16]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>414</td>
<td>403</td>
<td>24.0%</td>
<td>0.90</td>
<td>[0.76, 1.06]</td>
<td></td>
</tr>
</tbody>
</table>

Total events 177, 192
Heterogeneity Tau² = 0.01; Chi² = 7.93; df = 7 (P = 0.34); I² = 13%
Test for overall effect Z = 1.22 (P = 0.22)

Total (95% CI) 1392, 1377 100.0% 1.00 [0.93, 1.09]

Total events 718, 710
Heterogeneity Tau² = 0.00; Chi² = 10.53; df = 10 (P = 0.24); I² = 18%
Test for overall effect Z = 0.10 (P = 0.92)

The asymmetry coefficient for the main responder analysis was -1.07 (p = 0.09; see funnel plot in Figure 6). In univariable meta-regression analysis, there was a significant association between country of origin and response (trials from German-speaking countries favouring hypericum; p = 0.037). In the multivariable meta-regression analysis, none of the three tested predictors proved significant (R² = 0.24).
Analyses based on HAMD scores confirmed the findings of the responder analysis (see comparisons 5.1 to 5.5, and 5.8). Analyses of MADRS- and D-S-values are difficult to interpret, as only few trials reported these outcomes (see comparisons 5.6, 5.7, 5.9). There was no funnel plot asymmetry (0.30, p = 0.73). In the multivariable meta-regression analysis, trials with higher HAMD baseline values showed less favourable results (p = 0.010), while country of origin and precision had no significant influence (R² = 0.44).

Again, post-hoc analyses using available data for a variety of self-rating instruments from 10 trials comparing hypericum extracts and standard antidepressants confirmed analyses based on physician-rated outcomes. The pooled SMD was 0.01 (95%CI -0.13 to 0.15; I² = 43%; see comparison 5.10). The pooled SMDs in trials from German-speaking countries was -0.02 compared to 0.10 in trials from other countries (comparison 5.11).

2. Safety

Primary outcome

Patients allocated to hypericum extracts were less likely to drop out from studies due to adverse effects than patients allocated to older standard antidepressants (OR = 0.24; 95%CI 0.13 to 0.46; I² = 0%) or to SSRIs (OR = 0.53; 95%CI 0.34 to 0.83; I² = 0%; see comparison 6.1 and Figure 7).
Secondary outcomes

Attrition for any reason was significantly lower for hypericum extracts compared to older antidepressants (OR = 0.67; 95% CI 0.47 to 0.95; I² = 0%), but not compared to SSRIs (OR = 0.83; 95% CI 0.73 to 1.0; I² = 0%; see comparison 6.2).

The number of patients reporting side effects was significantly higher in patients receiving older standard antidepressants (OR = 0.39; 95% CI 0.30 to 0.50; I² = 67%); compared to SSRIs, the difference just missed significance (OR = 0.70; 95% CI 0.49 to 1.00; I² = 57%; see comparison 6.3).

Discussion

Summary of main results

Overall, the findings from newer trials seem to corroborate the evidence in favour of hypericum extracts. The available data suggest that the hypericum extracts tested in the included trials a) are superior to placebo in patients with major depression; b) are similarly effective as standard antidepressants; c) and have less side effects than standard antidepressants. There are two issues which complicate the interpretation of our findings: 1) While the influence of precision on study results in placebo-controlled trials is less pronounced in this updated version of our review compared to the previous version (Linde 2005a), results from more precise trials still show smaller effects over placebo than less precise trials. 2) Results from German-language countries are considerably more favourable for hypericum than trials from other countries.

Interpretation of the findings and limitations

For this update we excluded for the first time from our review all trials which were not restricted to patients with major depres-
sion. This does not mean that we believe that major depression is necessarily the only or best indication for hypericum extracts. Some authors argue that patients with signs of atypical depression might be particularly suited for treatment with hypericum extracts (Murck 2002, Murck 2005), the National Institutes of Health are currently funding a trial in patients with minor depression (www.clinicaltrials.gov, identifier NCT00048815), and the findings from older trials (some of which, however, seem methodologically questionable) not restricted to patients with major discussion were very positive (Linde 1998). As pointed out in the introduction, we now focus on major depression to make our review much more comparable to overviews on standard antidepressants, to have a more comparable set of studies for analysis, and also because almost all new trials of hypericum extracts are on this indication.

In spite of the tightened inclusion criteria, the findings of the placebo-controlled trials are still quite heterogeneous. In the trials a variety of hypericum extracts has been tested and daily doses cover a wide range. Differences in the interventions might contribute to some extent to the observed heterogeneity, but they do not seem to be a major factor. In three of four subgroup analyses of single extracts there was strong heterogeneity; for one extract the 95% confidence intervals of the two available trials (Gastpar 2006; Uebelhack 2004) did not even overlap, indicating that the results are hardly compatible. However, some of the factors leading to considerable heterogeneity between study findings could be identified. Considering country of origin, precision, and baseline depression severity of included patients explained 50 to 60 percent of the variance between trial results in comparisons with placebo and 20 to 40 percent in comparison with standard antidepressants. Nevertheless, it has to be stated that meta-regression analyses are (even if a priori defined) entirely of observational nature. Findings on the association of baseline depression severity and effect size estimates may be biased through structural dependence and regression to the mean, but this should be interpreted with caution (Higgins 2008). Furthermore, inferences drawn from a meta-regression analysis on aggregate data may differentiate from inferences drawn from a meta-regression analysis on individual data (Deeks 2006; e.g. ecological bias).

The finding that more precise placebo-controlled trials yielded less positive results than less precise trials could indicate publication bias (trials with positive results are more likely to be published than trials with negative results) or bias within studies (smaller trials with less rigorous methods yielding overoptimistic results). We cannot rule out, but doubt, that selective publication of overoptimistic results in small trials strongly influences our findings. There is some evidence that “negative” trials without demonstrable differences between extracts and placebo were published less often as full articles than trials with “positive” findings. Our extensive searches identified three “negative” trials that were only published as abstracts or theses. Two that were conducted in the early 1990s (these were included in earlier versions of this review Linde 1996; Linde 1998; Linde 2005a) involved patients without documented major depression (König 1993; Osterheider 1992), and one that was conducted in the late 1990s involved patients with major depression (Montgomery 2000). One positive trial included in our last update, but now excluded, was published as an abstract and as a chapter in a not widely available book (Winkel 2000). One comparably large, positive trial (Bracher 2001) has been published only in a short report as a supplement to a German medical newspaper. This trial is an example that sponsors or manufacturers of herbal medicines sometimes have very limited interest in a major publication if their trial includes a new aspect (in that case a once daily dosage), as there is no patent protection for herbal extracts and results can be exploited by competitors, too. We suspect that there are few additional relevant unpublished trials. Few manufacturers of hypericum extracts sponsor research trials, and the five manufacturers whose products were tested in most of the trials told us they had (with the exception of one smaller negative trial) no other unpublished trials that possibly met our inclusion criteria. Through personal communication we were informed that there are at least one or two unpublished negative trials on tea preparations of Hypericum. However, tea preparations are phytochemically very different from alcoholic extracts and have to be evaluated separately.

We found that the quality of the majority of trials was adequate, and we detected no systematic differences in design aspects known to be potential sources of bias. All trials were double-blind. Though adequacy of blinding was not formally assessed in most trials, achieving similarity between hypericum and placebo preparations is not particularly difficult. All trials were randomised, and most concealed allocation assignments by using consecutively numbered identical medication containers. Reported drop-out rates were low in the majority of trials. Investigators involved in older trials may have had less training and/or experience with diagnostic standards and rating scales for depressive symptoms (Shelton 2001), but this issue, if true, is likely to affect generalisation of findings rather than internal validity. Finally, though we found no systematic differences in major factors generally related to trial quality, our subjective judgement was that larger trials tended to be of better quality than smaller trials. The dosages of standard antidepressants were (with two exceptions) within the range recommended in current guidelines (e.g., Härter 2003), but at the lower limit.

Our finding that studies from German-speaking countries yielded more favourable results than trials performed elsewhere is difficult to interpret. As our analyses are partly data-driven, they must be considered cautiously. However, the consistency and extent of the observed association suggest that there are important differences in trials performed in different countries. One possibility is that studies performed in German-speaking countries with a long history of hypericum prescription by physicians enrolled slightly different patients in spite of similar inclusion criteria. With one exception (the extremely positive trial by Uebelhack 2004) performed in a re-
search clinic of a contract research organisation), all German studies recruited patients in private practices, while a number of trials from other countries were performed in academic research settings or hospital outpatient units. Depression with atypical or reversed vegetative features might be present more often in primary care outpatient populations (Murck 2005). The trend that trials with higher HAMD baseline values reported slightly less favourable results also suggests that effectiveness of hypericum extracts might differ between subgroups of depressive patients. While we did not systematically investigate this issue, it seems to us that the trials from countries other than Germany might be more often investigator-initiated. A closer link of trial planning, performance and analysis with manufacturer interests might influence study findings. This could result possibly in true bias, but also in conditions making a true positive outcome more likely. For example, for at least three trials (Kasper 2006; Schrader 1998; Uebelhack 2004) with large effects performed in German-speaking countries authors or sponsors reported in the publication or in personal communications that contact times and interaction with patients were limited to minimise placebo response rates. Increasing placebo group response rates due to the intensive care and monitoring in antidepressant trials are considered by some researchers as a potential reason for the problem to show specific effects (Posternak 2007). One could also speculate whether unblinding might lead German physicians (who often use hypericum extracts in their usual practice) to give more positive ratings and (the possibly more sceptic) colleagues from elsewhere to more negative ratings. However, as hypericum extracts have no characteristic side effects such a problem seems only relevant in comparisons with standard antidepressants.

Potential biases in the review process

The work for the first version of this review started in 1993 and three previous versions are available (Linde 1996; Linde 1998; Linde 2005a). During that period a large number of new trials became available, diagnostic classifications used for including patients into studies have changed and the quality of trials has improved. In parallel the methods of our review were adapted. The changes over time make it difficult to report our searches and their results in a consistent and transparent manner. The way how we approached authors/sponsors for obtaining missing information and the contents of inquiries were not fully systematic and have changed over time. This could imply that additional data necessary for some secondary analyses were obtained for a selected subset of studies. However, data for the main analysis were available for all or almost all trials, therefore, major biases seem highly unlikely. A potential source of bias in the responder analyses could be slightly variable responder definitions in the primary studies. Response according to the HAMD was either defined as at least 50% reduction, a HAMD score < 10 (or 11) after treatment, at least one or the combination of both. Whether these definitions were truly made a priori in each study could not be assessed. Decisions on the inclusion of subgroup analyses (for example, regarding precision or country effects) for updates were driven by findings in previous versions of the review. Therefore, these analyses must be interpreted with caution. Publication and small study bias have been discussed in the previous section.

Authors’ conclusions

Implications for practice

In older systematic reviews and meta-analyses of hypericum extracts (Ernst 1995; Linde 1996; Linde 1998; Kim 1999; Gaster 2000; Williams 2000) the findings of the included studies were mostly positive, but reviewers drew cautious conclusions due to methodological limitations. The quality of trials on average clearly improved over recent years. However, study findings became more often contradictory, and in the last version of our review larger trials restricted to patients with major depression showed only minimal effects over placebo (Linde 2005a). With the addition of several new, partly large trials, the cumulative evidence now suggests that hypericum extract have a modest effect over placebo in a similar range as standard antidepressants (Kirsch 2008; Turner 2008). The direct comparisons with older antidepressants and selective serotonine reuptake inhibitors seem to confirm this impression. The available clinical trials now also show that hypericum extracts have fewer side effects than both older antidepressants and selective serotonine reuptake inhibitors. This would imply that an attempt of treating mild to moderate major depression with one of the hypericum preparations positively tested in clinical trials is clearly justified. However, the differences in the findings from different countries make clear-cut recommendations difficult. The evidence for severe major depression is still insufficient to draw conclusions.

Many patients buy St John’s wort products from health food stores and might not disclose this to their physicians (Smith 2004). Such uncontrolled use is problematic as serious interactions can occur with a number of frequently used drugs (Ernst 1999; Hammerness 2003; Knüppel 2004; Whitten 2006). Therefore, physicians should regularly ask their patients about hypericum intake. However, it must be kept in mind that drug interactions are not a problem unique to hypericum extracts, but also common for standard antidepressants (Nieuwstraten 2006).

It has to be emphasised that the quality of hypericum preparations can differ considerably. The composition of a product depends on the raw plant material used, the extraction process, and the solvents. In consequence, the amounts of bioactive constituents in different products can vary enormously. A recent study has shown that a number of products available on the German market contain only minor amounts of bioactive constituents (Würglics 2003).
The hypericum extracts tested in clinical trials have to be considered high quality products. Results obtained with these extracts cannot be extrapolated directly to other products. In our meta-analysis, the type of extract did not contribute to the explanation of heterogeneity. This does not mean, however, that all products tested in the available trials are equally effective. Standardisation of a product on a defined component (for example, hyperforin or hypericin) does not resolve the problem, as currently the exact mechanism for the antidepressant effects of hypericum extracts is still unclear, and available research indicates that several components are relevant. The findings of this review most likely apply to products (using ethanol 50 to 60% or methanol 80% for extraction from dried plant material) with daily extract dosages of 500 to 1200 mg with a ratio of raw material to extract of 3:7:1.

Implications for research
There is a clear need to investigate the reasons for the differences in findings from trials originating from German-speaking countries and those from other countries. Multinational trials would seem desirable, but it is unlikely that there will be funding for such studies in the near future. Individual patient data meta-analysis of existing trials could be a possible tool to investigate predictors of treatment response in a more accurate manner. The authors will try to obtain such data from researchers and/or sponsors.

ACKNOWLEDGEMENTS
The authors would like to thank
- the researchers and sponsors who provided additional information on trials
- Eva Israel and Karsten Schmidt who helped with assessing eligibility and doing extractions
- Cindy Mulrow and Matthias Egger who were involved in previous versions of the review

REFERENCES

References to studies included in this review

Behnke 2002  [published data only]

Bjerkenstedt 2005  [published data only]

Bracher 2001  [published and unpublished data]

Brenner 2000  [published data only]

Brenner R, Madhusoodanan S, Pawlowska M. Efficacy of continuation treatment with hypericum perforatum in depression.
St John's wort for major depression (Review)

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Fava 2005 [published data only]

Gastpar 2005 [published and unpublished data]

Gastpar 2006 [published and unpublished data]

Harrer 1993 [published and unpublished data]

Harrer 1999 [published data only]

HDTSG 2002 [published and unpublished data]

Hängsen 1996 [published and unpublished data]


Kalb 2001 [published data only]

Kasper 2006 [published and unpublished data]

Laakmann 1998 [published data only]


Lecrubier 2002 [published and unpublished data]

Montgomery 2000 [published and unpublished data]

Moreno 2005 [published data only]

Philipp 1999 [published and unpublished data]

Schrader 1998 [published and unpublished data]

Schrader 2000 [published and unpublished data]

Shelton 2001 [published and unpublished data]

Szegedi 2005 [published and unpublished data]
Anghelescu IG, Kohnen R, Szegedi A, Klement S, Kieser M. Comparison of hypericum extract WS 5570 and paroxetine in ongoing treatment after recovery from an episode of moderate to


Uebelhack 2004 [published and unpublished data]


van Gurp 2002 [published and unpublished data]


Volz 2000 [published data only]


Vorbach 1997 [published and unpublished data]

Vorbach EU, Arnold KH, Hübner WD. Efficacy and tolerability of St John’s wort extract LI 160 versus imipramine in patients with severe depressive episodes according to ICD-10. *Pharmacopsychiatry* 1997;30(suppl.2):81–5.

Wheatley 1997 [published and unpublished data]


Witte 1995 [published and unpublished data]


Woelk 2000 [published data only]


References to studies excluded from this review

Agrawal 1994 [published data only]


Albertini 1986 [published data only]


Bendre 1980 [published data only]


Bergmann 1993 [published and unpublished data]


Bernhards 1993 [published data only]


Brockmöller 1997 [published data only]


Czekalla 1997 [published data only]


Dittmer 1992 [published data only]


Ditzler 1992 [published and unpublished data]


Gu 2001 [published data only]


Halama 1991 [published and unpublished data]


Harrer 1991 [published and unpublished data]


Herberg 1992 [published data only]


Hoffmann 1979 [published data only]

Hottenrott 1997  [published data only]

Häring 1996  [published and unpublished data]

Hübner 1993  [published and unpublished data]

Johnson 1992  [published data only]

Johnson 1993  [published data only]

König 1993  [published data only]

Lehrl 1993  [published and unpublished data]

Lenoir 1999  [published data only]
Lenoir S, Degenring FH, Saller R. A double-blind randomised trial to investigate three different concentrations of a standardised fresh plant extract obtained from the shoot tips of hypericum perforatum L. Phytomedizin 1999;6:141–6.

Li 2005  [published data only]

Maisenbacher 1995  [published data only]

Martinez 1993  [published data only]

Mo 2004  [published data only]

Osterheider 1992  [published and unpublished data]

Panijel 1985  [published data only]

Quandt 1993  [published and unpublished data]

Reh 1992  [published and unpublished data]
Sindrup 2000 (published and unpublished data)

Schmidt 1989 (published and unpublished data)

Schmidt 1993 (published and unpublished data)

Schmidt 1993b (published data only)

Schulz 1993 (published data only)

Sindrup 2000 (published data only)

Sommer 1994 (published and unpublished data)

Spielberger 1985 (published data only)

Staffeldt 1993 (published data only)

Steger 1985 (published and unpublished data)

Volz 2002 (published data only)

Vorbach 1994 (published and unpublished data)

Warnecke 1986 (published data only)

Werth 1989 (published data only)

Wiener 1991 (published data only)

Winkel 2000 (published data only)

Zeller 2000 (published data only)

Additional references
Caccia 2005

Candy 2008
252–60.


Reichling 2003

Smith 2004

Sterne 2001

Thompson 1999

Turner 2008

Ustun 2004

Volz 1997

von Zerssen 1996

Wheatley 1998

Whitten 2006

Williams 2000

Wilson 2002

Wurglics 2003

Wurglics 2006

References to other published versions of this review

Linde 1998

Linde 2005a

Linde 2005b

* Indicates the major publication for the study
### CHARACTERISTICS OF STUDIES

**Characteristics of included studies [ordered by study ID]**

**Behnke 2002**

| Methods | Concealment: unclear  
| Blinding: double-blind  
| Drop-outs/withdrawals: 6 of 35 in hypericum group, 3 of 35 in fluoxetine group  
| Jadad score: 1-1-1  
| IV score: 1-0-1-0.5-0.5-0.5 |
| Participants | Number of patients included/analyzed: 70/61  
| Demographics: 47 female, mean age 51 (treatment group) and 48 (control group)  
| Diagnosis: mild to moderate depression (ICD 10 F32.0 and 32.1)  
| Setting: multicenter trial from Denmark  
| Baseline: HAMD score treatment group 20.0 +/- 3.2, control group 20.7 +/- 2.9 |
| Interventions | Treatment: Hypericum extract (Calmigen) 2x1 coated tablet daily (300 mg extract) for 6 weeks  
| Control: Fluoxetine 2x20 mg for 6 weeks |
| Outcomes | Observation period: 6 weeks  
| Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction), Clinical Global Impression Index (CGI)  
| Patient-rated: von Zerssen Depression Scale, Clinical Global Impression Index (CGI), global assessment |
| Notes | Contract research organization which performed the study contacted for additional information but no answer was received |

<table>
<thead>
<tr>
<th>Risk of bias</th>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Unclear</td>
<td>B - Unclear</td>
<td></td>
</tr>
</tbody>
</table>

**Bjerkenstedt 2005**

| Methods | Concealment: consecutively numbered pharmacy  
| Blinding: double-blind (double-dummy)  
| Drop-outs/withdrawals: ITT analysis on 163 of 174 patients randomized  
| Jadad score: 1-2-0  
| IV score: 1-1-1-0.5-0.5-0.5 |
| Participants | Number of patients included/analyzed: 177 /163  
| Demographics: 129 female, mean age 50 years  
| Diagnosis: mild to moderate major depression (DSM-IV 296.31 or 32)  
| Setting: 15 practices (psychiatry, neurology, GP) in Sweden  
| Baseline: HAMD score hypericum group 24.9, fluoxetine group 23.8, placebo group 25.2 |
Interventions

Treatment: Hypericum extract LI 160 (Jarsin 300) 3x1 coated tablet (900 mg) for 6 weeks
Control 1: Fluoxetine 1x1 capsule (20 mg) for 6 weeks
Control 2: Placebo for 4 weeks, then randomized to hypericum or fluoxetine

Outcomes

Observation period: 6 weeks
Physician-rated: Hamilton Depression Scale (HAMD), Montgomery Asberg Depression Rating Scale (MADRS), Clinical Global Impression Index (CGI)

Notes

Due to the request of the ethical review board patients allocated to placebo had to be re-randomized to hypericum or fluoxetine after 4 weeks. Prior to publication of the main reference authors provided some additional information; after publication of the full report no further unpublished information was considered necessary

Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

Bracher 2001

Methods

Concealment: consecutively numbered pharmacy
Blinding: double-blind
Drop-outs/withdrawals: ITT analysis on 207 of 218 patients randomized
Jadad score: not performed (see notes)
IV score: not performed (see notes)

Participants

Number of patients included/analyzed: 218/207
Demographics: 167 female, mean age 44 years
Diagnosis: mild to moderate major depression (DSM-IV 296.31 or 32)
Setting: 17 practices (GP, various other) in Germany
Baseline: HAMD score hypericum group 19.7, placebo group 19.7

Interventions

Treatment: Hypericum extract HYP611 (Felis 650) 1x1 coated tablet (650 mg) for 6 weeks
Control: Placebo

Outcomes

Observation period: 6 weeks
Physician-rated: Hamilton Depression Scale (HAMD), Montgomery Asberg Depression Rating Scale (MADRS), Clinical Global Impression Index (CGI)
Patient-rated: von Zerssen Paranoid Depression Scale (PDS)

Notes

This trial was identified during the revision process. It is only published as a short report in a sponsored supplement to a nonscientific journal. The sponsor (Hexal AG, Holzkirchen, Germany) allowed a reviewer (KL) to extract detailed information from the full unpublished study report

Risk of bias
### Bracher 2001 (Continued)

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - adequate</td>
</tr>
</tbody>
</table>

### Brenner 2000

| Methods                       | Concealment: unclear  
                              | Blinding: double-blind (double-dummy)  
                              | Drop-outs/withdrawals: 7 of 15 hypericum patients, 3 of 15 sertraline patients  
                              | Jadad score: 1-2-0  
                              | IV score: 1-0-1-1-1-0 |
| Participants                  | Number of patients included/analyzed: 30/28  
                              | Demographics: 19 female, mean age 45 years  
                              | Diagnosis: mild to moderate depressive disorder  
                              | Setting: community hospital in USA  
                              | Baseline: HAMD score hypericum group 21.3 +/- 3.2, sertraline group 21.7 +/- 2.7 |
| Interventions                | Treatment: Hypericum extract LI 160 3x1 tablet (900 mg extract) for 7 weeks  
                              | Control: Sertraline 3x1 capsule (75 mg) for 7 weeks |
| Outcomes                     | Observation period: 7 weeks  
                              | Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction), Clinical Global Impression Index (CGI)  
                              | Patient-rated: von Zerssen Depression Scale (D-S) |
| Notes                        | Small study. High drop-out rate with more losses in the hypericum group. In spite of intention to treat analysis bias cannot be ruled out |

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Unclear</td>
<td>B - Unclear</td>
</tr>
</tbody>
</table>

### Fava 2005

| Methods                       | Concealment: unclear  
                              | Blinding: double-blind  
                              | Drop-outs/withdrawals: 18/45 in the hypericum group, 23 of 47 in the fluoxetine group and 22 of 43 in the placebo group  
                              | Jadad score: 1-1-0  
                              | IV score: 1-0-1-0.5-0.5-0 |
| Participants                  | Number of patients included/analyzed: 135/135  
                              | Demographics: 57% female, mean age 37 years  
                              | Diagnosis: mild to moderate major depression |
### Fava 2005

(Continued)

| Setting | 2 psychiatric outpatient departments in the US  
Baseline: HAMD score (17 items) 19.6 +/- 3.5 in the hypericum group, 19.9 +/- 2.9 in the fluoxetine and 19.6 +/- 3.1 in the placebo group |
| --- | --- |
| Interventions | Treatment: Hypericum extract LI 160 3x1 tablets (900 mg) for 12 weeks (+ 1 capsule placebo)  
Control 1: Fluoxetine 1x20 mg (capsule) for 12 weeks (+ 3x1 placebo tablets)  
Control 2: Placebo (1x1 capsule, 3x1 tablets) |
| Outcomes | Observation period: 12 weeks (+ 1 week placebo run-in)  
Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction), Clinical Global Impression Index (CGI)  
Patient-rated: Beck Depression Inventory (BDI) |
| Notes | High drop-out rate (intent to treat analysis); recruitment stopped before planned sample size was reached due to decision of the sponsor (lichtwer Pharma, Berlin). Only remission rates reported (and use instead of responder data for analyses) |

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Unclear</td>
<td>B - Unclear</td>
</tr>
</tbody>
</table>

### Gastpar 2005

| Methods | Concealment: numbered pharmacy  
Blinding: double-blind (double-dummy technique)  
Drop-out/withdrawals: 17 of 123 (hypericum), 19 of 118 (sertraline)  
Jadad score: 2-2-1  
IV score: 1-1-1-0.5-0.5-0.5 |
| --- | --- |
| Participants | Patients included/analyzed: 241/118 (in per protocol analysis)  
Demographics: 74% female, mean age 19 years  
Diagnosis: moderate depression (F32.1 or F33.1)  
Setting: 18 primary care physicians in Germany  
Baseline: HAMD values 22.0 +/- 1.1 vs. 22.1 +/- 1.1 |
| Interventions | Interventions: Hypericum extract STW3 1x1 tablet (612 mg extract) + 1 capsule placebo daily for 12 weeks  
Control: 1x1 tablet sertraline (50 mg) + 1 tablet placebo daily for 12 weeks  
After the 12 week-treatment phase there was an optional continuation phase |
| Outcomes | Observation period: 12 weeks for the main comparison + 12 weeks continuation phase  
Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = HAMD score < 10 or at least 50% reduction), Clinical Global Impression Index (CGI)  
Patient-rated: Adjective Mood Scale (BfS) |
The number of adverse effects in the hypericum group is much higher than in the sertraline group (189 vs. 112). For the number of patients reporting adverse effects the difference is less pronounced (74 vs. 60; 1 patient in the hypericum group had reported 26 adverse effects). In the hypericum group there was a higher number of infections (48 vs. 26; not observed in any other trial before). The number of adverse effects which were possibly causally related with the treatment was higher in the sertraline group (12 vs. 16). Also the intensity of side effects was more pronounced in the sertraline group.

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

### Gastpar 2006

#### Methods
- Concealment: numbered pharmacy
- Blinding: double-blind (double-dummy technique)
- Drop-outs/withdrawals: 6 of 131 (hypericum), 6 of 127 (citalopram), 8 of 130 (placebo)
- Jadad score: 2-2-1
- IV score: 1-1-1-0.5-0.5-1

#### Participants
- Patients included/analyzed: 388/388
- Demographics: 67% female, mean age 50 years
- Diagnosis: moderate depression (ICD-10 F32.1 or F33.1)
- Setting: 21 general practitioners and internists in Germany
- Baseline: HAMD score hypericum group 21.9 +/- 1.2, citalopram group 21.8 +/- 1.2, placebo group 22.0 +/- 1.2

#### Interventions
- Treatment: Hypericum extract STW3-VI 1x1 tablet (900 mg extract) for 6 weeks
- Control 1: Placebo for 6 weeks
- Control 2: Citalopram 1x1 tablet 20 mg for 6 weeks

#### Outcomes
- Observation period: 6 weeks
- Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = HAMD score < 10 or at least 50% reduction), Clinical Global Impression Index (CGI)
- Patient-rated: Adjective Mood Scale (BfS)

#### Notes
- Additional information provided by first author and sponsor

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>
Harrer 1993

**Methods**
- Concealment: consecutively numbered pharmacy
- Blinding: Double-blind
- Drop-outs/withdrawals: 7 of 51 in hypericum group, 9 of 51 in maprotiline group
- Jadad score: 2-2-1
- IV score: 1-1-0.5-0.5-0.5

**Participants**
- Number of patients included/analyzed: 102 /86
- Demographics: 73 female, mean age 44 years in the Hypericum group and 48 years in the maprotiline group
- Diagnosis: single, moderately severe depressive episode (ICD 10 F32.1) (according to information from the sponsor patients met DSM-III-R criteria for major depression); Setting: 6 practices in Germany (neurology/psychiatry)
- Baseline: HAMD scores 20.5 +/- 3.7 in hypericum group and 21.5 +/- 3.9 in maprotiline group

**Interventions**
- Treatment: Hypericum extract LI 160 (Jarsin 300) 3x1 coated tablet (900 mg extract) daily for 4 weeks
- Control: Maprotiline 3x1 coated tablet (75 mg) daily for 4 weeks

**Outcomes**
- Observation period 4 weeks
- Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction or score x 10), Clinical Global Impression Index (CGI), global assessment
- Patient rated: Depression Scale von Zerssen D-S, global assessment

**Notes**
- Additional information provided from sponsor (Lichtwer, Berlin, Germany)

**Risk of bias**

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

Harrer 1999

**Methods**
- Concealment: unclear
- Blinding: double-blinding
- Drop-outs/withdrawals: 8 of 77 hypericum patients and 16 of 84 fluoxetine patients
- Jadad score: 1-1-1
- IV score: 1-0-1-0.5-0.5-0.5

**Participants**
- Number of patients included/analyzed: 161/149
- Demographics: 129 female, mean age 69 years
- Diagnosis: mild to moderate depression (ICD 10 F32.0 or F32.1)
- Setting: 17 practices in Germany
- Baseline: HAMD score hypericum group 16.6, fluoxetine group 17.2

**Interventions**
- Treatment: Hypericum extract LoHyp-57 2x2 coated tablets (800 mg) for 6 weeks
- Control: Fluoxetine 2x2 coated tablets (20 mg) for 6 weeks
Outcomes
Observation period: 6 weeks
Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction or score < 11), Clinical Global Impression Index (CGI)
Patient-rated: Self-Rating Scale for Depression (SDS), Fragebogen Alltagsleben (German quality of life questionnaire)

Notes
Reporting of results partly insufficient

Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Unclear</td>
<td>B - Unclear</td>
</tr>
</tbody>
</table>

HDTSG 2002

Methods
Concealment: central telephone randomization
Blinding: double-blind (double-dummy)
Drop-outs/withdrawals: 31 of 113 in hypericum group, 32 of 116 in placebo group, and 32 of 111 in sertraline group
Jadad score: 2-2-1
IV score: 1-1-1-0.5-0.5-0.5

Participants
Number of patients included/analyzed: 340/340
Demographics: 66% female, mean age 43 years
Diagnosis: major depression (DSM-IV)
Setting: 12 academic and community psychiatric research clinics in the US
Baseline: HAMD scores 23.1 +/- 2.7 (hypericum), 22.7 +/- 2.7 (placebo), 22.5 +/- 2.5 (sertraline)

Interventions
Treatment: Hypericum LI 160 extract 900 to 1500 mg for 8 weeks
Control 1: Placebo for 8 weeks
Control 2: Sertraline 50 to 100 mg for 8 weeks

Outcomes
Observation period: 1 weeks run-in, 8 weeks treatment, 18 weeks follow-up for responders
Physician-rated: Hamilton Depression Scale (HAMD, response = at least 50% reduction or score < 10), Clinical Global Impression Index (CGI), Global Assessment of Functioning (GAF)
Patient-rated: Beck Depression Inventory (BDI), Sheehan Disability Scale (SDS)

Notes
Trial funded by public institution (NIH). Large proportion of patients with chronic depression. Some unblinding detected (sole trial which reported a check of blinding). Authors provided additional data

Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>
Hänsgen 1996

Methods
Concealment: numbered pharmacy
Blinding: double-blind
Drop-outs/withdrawals: 3 of 54 in hypericum group, 4 of 54 in placebo group
Jadad score: 2-2-1
IV score: 1-1-1-1-1-0.5

Participants
Number of patients included/analyzed: 108/101
Demographics: 66 female, mean age 52 years
Diagnosis: major depression (DSM-III-R, HAMD score > 15)
Setting: 17 practices in Germany (neurologists/psychiatrist, general practitioners)
Baseline: HAMD score 21.8 +/- 2.8 (hypericum), 20.4 +/- 3.4 (placebo)

Interventions
Treatment: Hypericum extract LI160 (Jarsin 300) 3x1 coated tablet daily (900 mg extract) for 4 weeks
Control: Placebo
For further 2 weeks both groups received Hypericum

Outcomes
Observation period: 4 weeks
Physician-rated: Hamilton Depression Scale (HAMD, 21 items; response = at least 50% reduction or score < 10), Clinical Global Impression Index (CGI) after 2 and 4 weeks
Patient-rated: Depression Scale von Zerssen (D-S), complaints check list (BEB) after 2 and 4 weeks

Notes
Additional information provided by author and sponsor (Lichtwer, Berlin, Germany). This trial was first published in 1993 (in German, 1994 in English) with 72 patients and re-published with 108 patients in 1996 (without refering to the earlier publications)

Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

Kalb 2001

Methods
Concealment: numbered pharmacy
Blinding: double-blind
Drop-outs/withdrawals: none
Jadad score: 2-2-1
IV score: 1-1-1-1-1-1

Participants
Patients included/analyzed: 72/72
Demographics: 48 female, mean age 48 years
Diagnosis: mild to moderate depression (DSM-IV 296.21/31/22/32)
Setting: 11 practices (psychiatry, internal medicine, GP) in Germany
Baseline: HAMD score 19.7 +/- 3.4 hypericum group, 20.1 +/- 2.6 placebo group

Interventions
Treatment: Hypericum extract WS 5572 (Neuroplant)3x1 coated tablet (900 mg) for 6 weeks
Control: Placebo for 6 weeks
### Kalb 2001 (Continued)

| Outcomes | Observation period: 3-7 days run-in, 6 weeks treatment  
|          | Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction), Clinical Global Impression Index (CGI)  
|          | Patient-rated: von Zerssen Depression Scale (D-S), Patient’s Global Assessment Scale (GPA)  
| Notes | Trial with adaptive design stopped early due to significant superiority at preplanned interim analysis  

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

### Kasper 2006

| Methods | Concealment: numbered pharmacy  
|         | Blinding: double-blind  
|         | Drop-out/withdrawals: 12 of 123 (hypericum 600 mg), 19 of 127 (hypericum 1200 mg), 8 of 82 (placebo)  
|         | Jadad score: 2-2-1  
|         | IV score: 1-1-0.5-1-1-1 |
| Participants | Patientns included/analyzed: 332/324  
|              | Demographics: 63% female, mean age 46 years  
|              | Diagnosis: mild or moderate, single or recurrent, major depressive episode (DSM IV 296.21/22, 296.31/32)  
|              | Setting: 11 psychiatric and 5 GP practices in Germany  
|              | Baseline: HAMD score hypericum 600 mg group 22.8 +/- 3.3, hypericum 1200 mg group 22.6 +/- 3.8, placebo group 23.6 +/- 4.2 |
| Interventions | Treatment 1: 1 tablet hypericum extract (600 mg) daily + 1 placebo tablet for 6 weeks  
|               | Treatment 2: 2x1 tablet hypericum extract (total daily extract dosage 1200 mg) daily for 6 weeks  
|               | Control: 2x1 tablet placebo daily for 6 weeks |
| Outcomes | Observation period: 6 weeks  
|          | Physician-rated: Hamilton depression Scale (HAMD, 17 items; response = at least 50% score reduction), Clinical Global Impression Index (CGI), Montgomery-Asberg Depression Scale (MADRS)  
|          | Patient-rated: Beck Depression Inventory (BDI), quality of life (SF-36)  
| Notes | Randomization in 3:3:2 ratio  
|       | Additional information provided by sponsor  

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>
### Laakmann 1998

| Methods                        | Concealment: unclear  
|                               | Blinding: double-blind  
|                               | Drop-outs/withdrawals: 3 in group hypericum 1, 1 in group hypericum 2, 4 in placebo group  
|                               | Jadad score: 2-2-1  
|                               | IV score: 1-0-1-1-1-1  
| Participants                   | Patients included/analyzed: 147/147  
|                               | Demographics: 117 female, mean age 49 years  
|                               | Diagnosis: mild to moderate depression (DSM-IV)  
|                               | Setting: 11 practices in Germany  
|                               | Baseline: HAMD score 20.9 +/- 3.1 (hypericum 1), 20.3 +/- 2.7 (hypericum 2), 21.2 +/- 3.3 (placebo)  
| Interventions                  | Treatment 1: Hypericum extract WS 5572 (5% hyperforin)3x1 coated tablet (900 mg) for 6 weeks  
|                               | Treatment 2: Hypericum extract WS 5573 (0.5% hyperforin)3x1 coated tablet (900 mg) for 6 weeks  
|                               | Control: Placebo for 6 weeks  
| Outcomes                       | Observation period: 6 weeks  
|                               | Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction), Clinical Global Impression Index (CGI)  
|                               | Patient-rated: von Zerssen Depression Scale (D-S), global assessment  
| Notes                          | Includes a group receiving a second hypericum extract with very low hyperforin content (not included in meta-analysis as this tests an extract which was not marketed)  

#### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

### Lecrubier 2002

| Methods                        | Concealment: numbered pharmacy  
|                               | Blinding: double-blind  
|                               | Drop-outs/withdrawals: 18 of 186 (hypericum), 25 of 189 placebo  
|                               | Jadad score: 1-2-0  
|                               | IV score: 1-0-1-1-1-0.5  
| Participants                   | Patients included/analyzed: 375/375  
|                               | Demographics: 297 female, mean age 41 years  
|                               | Diagnosis: mild to moderate depression (DSM-IV 296.21/22/31/32)  
|                               | Setting: 26 psychiatric centers and practices in France  
|                               | Baseline: HAMD score hypericum group 21.9 +/- 1.7, placebo group 21.9 +/- 1.7  
| Interventions                  | Treatment: Hypericum extract WS 5570 3x1 tablet (900 mg) for 6 weeks  
|                               | Control: placebo for 6 weeks  


Lecrubier 2002  (Continued)

| Outcomes | Observation period: 3-7 days run-in, 6 weeks treatment  
Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction), Montgomery-Asberg Depression Rating Scale (MADRS), Clinical Global Impression Index (CGI)  
Patient-rated: Symptom Check List (SCL-58) |
|---|---|
| Notes | Trial with preplanned interim analysis with 169 patients (no significant difference)  
Additional information provided by sponsor (Schwabe, Karlsruhe) |

**Risk of bias**

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

Montgomery 2000

| Methods | Concealment: unclear  
Blinding: double-blind  
Drop-outs/withdrawals: unclear  
Jadad score: 1-2-0  
IV score: 1-0-0-1-1-0.5 |
|---|---|
| Participants | Patients included/analyzed: 247/?  
Demographics: 183 female, mean age 43 years  
Diagnosis: mild to moderate depression (DSM-IV 296.2x/3x)  
Setting: 18 GPs and psychiatric outpatient clinics in the UK  
Baseline: HAMD about 22 in both groups (data extrapolated from figure in poster) |
| Interventions | Treatment: Hypericum extract LI 160 3x1 coated tablet (900 mg) for 12 weeks  
Control: Placebo for 12 weeks |
| Outcomes | Observation period: 3 to 7 days run-in, 12 weeks treatment  
Physician-rated: Hamilton depression Scale (HAMD, 17 items; response = at least 50% reduction or score < 10), Montgomery-Asberg Depression Rating Scale (MADRS), Clinical Global Impression Index (CGI)  
Patient-rated: None mentioned |
| Notes | Available as congress abstract and poster handout only. Additional information from sponsor (Lichtwer, Berlin) |

**Risk of bias**

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Unclear</td>
<td>B - Unclear</td>
</tr>
</tbody>
</table>
### Moreno 2005

**Methods**

- Concealment: unclear
- Blinding: double-blind (double-dummy technique)
- Drop-outs/withdrawals: 19 of 72 (no details for single groups reported)
- Jadad score: 1-1-0
- IV score: 1-0-0-0.5-0.5-0

**Participants**

- Number of patients included/analyzed: 72/66
- Demographics: 83% female, mean age 40 years
- Diagnosis: mild to moderate major depression (DSM-IV criteria)
- Setting: Institute of Psychiatry, University Sao Paulo, Brazil
- Baseline: extrapolated from figure HAMD score hypericum group 15, placebo 17, fluoxetine, 15

**Interventions**

- Treatment: Hypericum extract Iperisan 3x1 (900 mg) daily for 8 weeks
- Control 1: Placebo
- Control 2: Fluoxetine 20 mg/day for 8 weeks (1 capsule fluoxetine and 2 capsules placebo per day)

**Outcomes**

- Observation period: 1-week prerandomization wash-out, 8 weeks post randomization
- Physician-rated: Hamilton Depression Scale (HAMD, 21 items, response = at least 50% reduction, remission < 8 on the 17 item scale), Montgomery-Asberg Depression rating Scale (MADRS), Clinical Global Impression Index (CGI)
- Patient-rated: none

**Notes**

- Publication misses important details. Authors inquired for additional information but no response received

### Philipp 1999

**Methods**

- Concealment: numbered pharmacy
- Blinding: Double-blind
- Drop-outs/withdrawals: 13 of 106 (hypericum), 11 of 110 (imipramine), 9 of 47 (placebo)
- Jadad score: 2-2-1
- IV score: 1-1-1-0.5-0.5-0.5

**Participants**

- Patients included/analyzed: 263/251
- Demographics: 197 female, mean age 47 years
- Diagnosis: moderate depression (ICD-10 F32.1/F32.2)
- Setting: 18 GPs in Germany
- Baseline: HAMD score 22.7 +/- 4.2 (hypericum), 22.2 +/- 4.2 (imipramine), 22.7 +/- 4.0 (placebo)

**Interventions**

- Treatment: Hypericum extract STEI 300 3x1 capsule (1050 mg) for 8 weeks
- Control 1: Imipramine 100 mg for 8 weeks
- Control 2: Placebo for 8 weeks

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Unclear</td>
<td>B - Unclear</td>
</tr>
</tbody>
</table>

St John's wort for major depression (Review)
### Philipp 1999

**Outcomes**
- Observation period: 1 week screening, 8 weeks treatment
  - Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction), Hamilton Anxiety Scale (HAMA), Clinical Global Impression Scale (CGI)
  - Patient-rated: Zung Depression Scale, Quality of life (SF-36)

**Notes**
- Additional information from sponsor (Steiner, Berlin). Primary outcome comparison with placebo after 6 weeks, with imipramine after 8 weeks

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

### Schrader 1998

**Methods**
- Concealment: numbered pharmacy
- Blinding: double-blind
- Drop-outs/withdrawals: 1 of 81 (hypericum), 2 of 81 (placebo)
- Jadad score: 2-2-1
- IV score: 1-0-0-1-1-1

**Participants**
- Patients included/analyzed: 162/159
- Demographics: 108 female, mean age 47 (hypericum) vs. 39 (placebo)
- Diagnosis: mild to moderate depression (F32.0/32.1)
- Setting: 16 private practices in Germany
- Baseline: HAMD values 20.1 +/- 2.8 (hypericum) vs. 18.7 +/- 3.5

**Interventions**
- Treatment: 2x1 coated tablet (500 mg extract) Hypericum extract ZE117 daily for 6 weeks
- Control: Placebo

**Outcomes**
- Observation period: 6 weeks
  - Physician-rated: Hamilton Depression Scale (HAMD, 21 items; response = at least 50% reduction or score < 10), Clinical Global Impression Index (CGI)
  - Patient-rated: visual analogue scale

**Notes**
- Additional information provided by author and sponsor (Zeller AG, Romanshorn, Switzerland). Baseline HAMD score lower in placebo group. Low response rate in placebo group.
- In meta-analyses HAMD values were calculated for the 17-item version (data provided by author)

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>
### Schrader 2000

**Methods**
- Concealment: numbered pharmacy
- Blinding: double-blind (double-dummy)
- Drop-outs/withdrawals: 1 of 126 (hypericum), 1 of 113 (fluoxetine)
- Jadad score: 1-2-1
- IV score: 1-0-1-0.5-0.5-1

**Participants**
- Patients included/analyzed: 230/228
- Demographics: 157 female, mean age 46 years
- Diagnosis: mild to moderate depression (ICD-10 F32.0/32.1)
- Setting: 7 practices (internal medicine) in Germany
- Baseline: HAMD values 19.6 +/- 3.1 (hypericum), 19.5 +/- 2.4 (fluoxetine)

**Interventions**
- Treatment: Hypericum extract Ze 117 2x1 coated tablet (500 mg) for 6 weeks
- Control: Fluoxetine 1x1 capsule (20 mg) for 6 weeks

**Outcomes**
- Observation period: 6 weeks
- Physician-rated: Hamilton Depression Scale (HAMD, 21 items; response = at least 50% reduction or score < 10), Clinical Global Impression (CGI)
- Physician-rated: visual analogue scale

**Notes**
- Additional information (particularly responder rates and 17-item HAMD scores) provided by author and sponsor (Zeller AG, Romanshorn, Switzerland)

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

### Shelton 2001

**Methods**
- Concealment: numbered pharmacy
- Blinding: double-blind
- Drop-outs/withdrawals: 15 of 98 (hypericum), 13 of 102 (placebo)
- Jadad score: 2-2-1
- IV score: 1-1-1-1-1-1

**Participants**
- Patients included/analyzed: 200/195
- Demographics: 64% female, mean age 42 years
- Diagnosis: major depression (DSM-IV)
- Setting: 11 academic medical centers in the USA
- Baseline: mean HAMD values above 22 in both groups

**Interventions**
- Treatment: Hypericum extract LI 160 3x1 to 4x1 tablet (900 to 1200 mg) for 8 weeks
- Control: placebo for 8 weeks

**Outcomes**
- Observation period: 1 week run-in, 8 weeks treatment
- Physician-rated: Hamilton Depression Scale (HAMD, 17 items, at least 50% reduction), Clinical Global
<table>
<thead>
<tr>
<th>Impression Index (CGI), Hamilton Anxiety Scale (HAMA) Patient-rated: Beck Depression Inventory (BDI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notes</td>
</tr>
</tbody>
</table>

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

### Szegedi 2005

**Methods**
- Allocation concealment: numbered pharmacy
- Blinding: double-blind (double dummy technique)
- Drop-out/withdrawals: 17 of 125 in hypericum group, 29 of 126 in the paroxetine group
- Jadad score: 2-2-1
- IV score: 1-1-1-0.5-0.5-0.5

**Participants**
- Number of patients included/analyzed: 251/244
- Demographics: 69% female, mean age 47 years
- Diagnosis: moderate or severe unipolar major depression (DSM IV 296.22/23, 296.32/33)
- Setting: 21 Psychiatric primary care centers in Germany
- Baseline: HAMD score 25.5 +/- 2.7 (hypericum), 25.5 +/- 2.9 (paroxetine)

**Interventions**
- Treatment: 3x1 tablet (900 mg daily) hypericum extract WS 5570 daily for 6 weeks (in patients whose depression score was not improved by at least 20% 3x 600 mg were used after 2 weeks)
- Control: 1x1 tablet (20 mg) paroxetine daily for 6 weeks (in patients whose depression score was not improved by at least 20% 40 mg were used after 2 weeks) packed in capsules
- For each drug an identically matched placebo was available

**Outcomes**
- Observation period: 6 weeks (+ 16 weeks continuation phase)
- Physician-rated: Hamilton Depression Scale (HAMD, 17 items, at least 50% reduction), Clinical Global Impression (CGI), Montgomery-Asberg Depression Rating Scale (MADRS)
- Patient-rated: Beck Depression Inventory (BDI)

**Notes**
- Additional information provided by sponsor

### Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>
### Uebelhack 2004

| **Methods** | Concealment: consecutively numbered pharmacy  
Blinding: double-blind  
Drop-out/withdrawals: none  
Jadad score: 2-2-1  
IV score: 1-0-1-1-1-1 |
|---|---|
| **Participants** | Number of patients included/analyzed: 140/140  
Demographics: 67% female, mean age 45 years  
Diagnosis: moderate depressive disorder (ICD-10 F32.1 or F33.1)  
Setting: clinical trial center in Germany  
Baseline: 22.8 +/- 1.1 in hypericum group, 22.8 +/- 1.2 in placebo group |
| **Interventions** | Treatment: 1x1 tablet (900 mg) STW3-VI daily for 6 weeks  
Control: Placebo |
| **Outcomes** | Observation period: 6 weeks  
Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = HAMD score < 10 or at least 50% reduction), Clinical Global Impression Index (CGI)  
Patient-rated: Adjective Mood Scale (BfS) |
| **Notes** | Trial performed in a single clinical trial unit with short recruitment period and low placebo response. Additional information from sponsor |

**Risk of bias**

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

### van Gurp 2002

| **Methods** | Concealment: independent pharmacist  
Blinding: double-blind  
Drop-outs/withdrawals: 16 of 45 (hypericum), 17 of 45 (sertraline)  
Jadad score: 2-2-1  
IV score: 1-1-1-0.5-0.5-0 |
|---|---|
| **Participants** | Patients included/analyzed: 90/87  
Demographics: 61% female, mean age 40 years  
Diagnosis: major depression (DSM IV)  
Setting: 12 community based offices for family medicine in Canada  
Baseline: mean HAMD values 18.8 +/- 3.6 (hypericum group), 19.7 +/- 3.5 (sertraline group) |
| **Interventions** | Treatment: 3x1 to 3x2 (in case of insufficient response at 4 weeks) capsules (900-1800 mg) Hypericum extract daily for 12 weeks  
Control: 3x1 to 3x2 (in case of insufficient response at 4 weeks) capsules Sertraline (50 to 100 mg) daily for 12 weeks  
In 9 patients of each group the dosage was doubled during the trial |
van Gurp 2002  \textit{(Continued)}

| Outcomes | Observation period: 6 weeks  
|          | Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction or score < 10)  
|          | Patient-rated: Beck Depression Inventory (BDI)  
| Notes | Additional information provided from author. High drop-out rate  

**Risk of bias**

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>

Volz 2000

| Methods | Concealment: numbered pharmacy  
|         | Blinding: double-blind  
|         | Drop-outs/withdrawals: 2/71 (hypericum), 5/71 (placebo)  
|         | Jadad score: 1-2-0  
|         | IV score: 1-1-1-1-0.5  
| Participants | Patients included/analyzed: 142/140  
|         | Demographics: 81% female, mean age 47 years  
|         | Diagnosis: mild or moderate episode of a major depression (DSM-IV)  
|         | Setting: 17 practices for psychiatry, neurology, internal medicine and GP in Germany  
|         | Baseline: HAMD 21.0 +/- 2.0 (hypericum group), 20.7 +/- 1.9 (placebo group)  
| Interventions | Treatment: 2x1 capsule (500 mg) Hypericum extract D-0496 daily for 6 weeks  
|         | Control: Placebo  
| Outcomes | Observation period: 1 week placebo run in, 6 weeks treatment  
|         | Physician-rated: Hamilton Depression Scale (HAMD, 21 items; response = at least 50% reduction), Clinical Global Impression Index (CGI)  
|         | Patient-rated: von Zerssen Depression Scale (D-S)  
| Notes | Manufacturer/sponsor no longer existing. Author could provide only minimal additional information  

**Risk of bias**

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Yes</td>
<td>A - Adequate</td>
</tr>
</tbody>
</table>
### Vorbach 1997

**Methods**
- Concealment: numbered pharmacy
- Blinding: double-blind
- Drop-outs/withdrawals: 9 of 107 in hypericum group, 14 of 102 in imipramine group
- Jadad score: 2-2-1, IV score: 1-1-0.5-0.5-0.5-0.5

**Participants**
- Number of patients included/analyzed: 209/209 (186 per protocol)
- Demographics: 154 female, mean age 49 years
- Diagnosis: severe episode of a major depression
- Setting: 20 psychiatric practices
- Baseline: HAMD score 25.3 +/- 4.7 (hypericum group), 26.1 +/- 4.8 (imipramine group)

**Interventions**
- Treatment: 3x2 coated tablets (1800 mg extract) Hypericum extract LI 160 daily for 6 weeks
- Control: 3x2 coated tablets imipramine (150 mg) daily for 6 weeks

**Outcomes**
- Observation period: 6 weeks
- Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction), Clinical Global Impression Index (CGI)
- Patient-rated: Depression-Scale (D-S von Zerssen)

**Notes**
- Additional information provided from sponsor

### Wheatley 1997

**Methods**
- Concealment: numbered pharmacy
- Blinding: double-blind
- Drop-outs/withdrawals: 20 of 87 in hypericum group, 24 of 78 amitriptyline group
- Jadad score: 2-2-0, IV score: 1-1-0.5-0.5-0.5-0.5

**Participants**
- Number of patients included/analyzed: 165/156
- Demographics: 126 female, mean age 40 years
- Diagnosis: major depression (DSM-IV)
- Setting: 18 general practices and one hospital outpatient clinic in the UK
- Baseline: HAMD score 20.6 +/- 2.1 (hypericum group), 20.8 +/- 2.3 (amitriptyline group)

**Interventions**
- Treatment: 3x1 coated tablet (900 mg extract) Hypericum extract LI 160 daily for 6 weeks
- Control: 3x1 coated tablet amitriptyline (75 mg) daily for 6 weeks

**Outcomes**
- Observation period: 6 weeks
- Physician-rated: Hamilton Depression Scale (HAMD, 17 items; response = at least 50% reduction or score < 10), Montgomery-Asberg Reating Scale for Depression (MADRS), Clinical Global Impression Index (CGI)
Wheatley 1997  (Continued)

| Notes | Some additional information provided from sponsor. HAMD mean values extrapolated from figure |

| **Risk of bias** |
|------------------|----------------------------------|
| **Item** | **Authors' judgement** | **Description** |
| Allocation concealment? | Yes | A - Adequate |

Witte 1995

<table>
<thead>
<tr>
<th>Methods</th>
<th>Concealment: numbered pharmacy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blinding: double-blind</td>
</tr>
<tr>
<td></td>
<td>Drop-outs/withdrawals: 1 of 48 in hypericum group, 1 of 49 in placebo group</td>
</tr>
<tr>
<td></td>
<td>Jadad score: 1-2-1</td>
</tr>
<tr>
<td></td>
<td>IV score: 1-0-0.5-1.1-0.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participants</th>
<th>Number of patients included/analyzed: 97/95</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Demographics: 64 female, mean age 43 years</td>
</tr>
<tr>
<td></td>
<td>Diagnosis: depression (ICD-10 F32.1)</td>
</tr>
<tr>
<td></td>
<td>Setting: 5 general practitioners in Germany</td>
</tr>
<tr>
<td></td>
<td>Baseline: HAMD score 24.6 +/- 5.4 (hypericum group), 22.7 +/- 4.4 (placebo group)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interventions</th>
<th>Treatment: 2x1 capsules (200 to 240 mg) Hypericum extract (Psychotolin forte) daily for 6 weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control: placebo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Observation period: 6 weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Physician-rated: Hamilton Depression Scale (HAMD, 21 items; response = at least 50% reduction or score &lt; 10), Clinical Global Impression Index (CGI)</td>
</tr>
<tr>
<td></td>
<td>Patient-rated: Depression Scale (D-S von Zerssen), State Trait Anxiety Inventory (STAI)</td>
</tr>
</tbody>
</table>

| Notes | Additional information provided by author |

| **Risk of bias** |
|------------------|----------------------------------|
| **Item** | **Authors' judgement** | **Description** |
| Allocation concealment? | Yes | A - Adequate |
Methods

Concealment: unclear
Blinding: Double-blind (double dummy)
Drop-outs/withdrawals: 15/157 (hypericum), 32 of 167 (imipramine)
Jadad score: 2-2-0
IV score: 1-0-1-0.5-0.5-0

Participants

Patients included/analyzed: 324/324
Demographics: 71% female, mean age 46 years
Diagnosis: mild to moderate depression (ICD10 F32.0 or 1, F33.0 or 1)
Setting: 40 practices for psychiatry, internal medicine, GP in Germany
Baseline: HAMD 22.4 +/-3.4 (hypericum), 22.1 +/- 2.9 (imipramine)

Interventions

Treatment: 2x1 coated tablet (500 mg extract) Hypericum extract ZE 117 daily for 6 weeks
Control: Imipramine 150 mg daily for 6 weeks

Outcomes

Observation period: 6 weeks
Physician-rated: Hamilton Depression Scale (HAMD, 17 items; at least 50% reduction), Clinical Global Impression Index (CGI)
Patient-rated: global assessment

Notes

Some additional information by manufacturer of the extract

Risk of bias

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation concealment?</td>
<td>Unclear</td>
<td>B - Unclear</td>
</tr>
</tbody>
</table>

Characteristics of excluded studies  [ordered by study ID]

<table>
<thead>
<tr>
<th>Study</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agrawal 1994</td>
<td>Not major depression (Study in patients with fatigue and minor depression)</td>
</tr>
<tr>
<td>Albertini 1986</td>
<td>Not depression (RCT of hypericum in homeopathic preparation for dental neuralgia)</td>
</tr>
<tr>
<td>Bendre 1980</td>
<td>Not depression (CCT of homeopathic preparations of hypericum and arnica in dental practice)</td>
</tr>
<tr>
<td>Bergmann 1993</td>
<td>Inadequate dosage of amitriptyline (30 mg daily) - RCT included in previous versions of this review</td>
</tr>
<tr>
<td>Bernhadt 1993</td>
<td>RCT comparing two doses of hypericum for depression</td>
</tr>
<tr>
<td>Brockmöller 1997</td>
<td>RCT on pharmacokinetics and photosensitivity in healthy volunteers</td>
</tr>
<tr>
<td>Czekalla 1997</td>
<td>Report of results on ECG analyses undertaken in the Vorbach 1997 study</td>
</tr>
<tr>
<td>Reference</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Dittmer 1992</td>
<td>Only a minority of patients suffered from depression (RCT of Hypericum for &quot;psychovegetative complaints&quot;)</td>
</tr>
<tr>
<td>Ditzler 1992</td>
<td>Placebo-controlled randomized controlled trial in depressed patients using a fixed combination of several herbs including hypericum</td>
</tr>
<tr>
<td>Gu 2001</td>
<td>RCT comparing 900 mg of a hypericum extract with fluoxetine in patients with depression according to the Chinese classification CCMD-2. As it is not clear whether all included patients meet criteria for major depression the study was excluded. Response rates were similar in both groups</td>
</tr>
<tr>
<td>Halama 1991</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Harrer 1991</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Herberg 1992</td>
<td>RCT comparing hypericum and a combination of Hypericum and Valerian on concentration, reaction etc. in healthy volunteers</td>
</tr>
<tr>
<td>Hoffmann 1979</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Hottenrott 1997</td>
<td>Not depression (RCT of a combination of hypericum and vitamine E to enhance performance of athletes)</td>
</tr>
<tr>
<td>Haring 1996</td>
<td>Placebo-controlled RCT of hypericum to prevent reactive depression in 28 patients with solid tumors undergoing chemotherapy</td>
</tr>
<tr>
<td>Hübner 1993</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Johnson 1992</td>
<td>RCT on neurophysiological effects of hypericum in healthy volunteers</td>
</tr>
<tr>
<td>Johnson 1993</td>
<td>RCT comparing neurophysiological effects of hypericum and maprotiline in healthy volunteers</td>
</tr>
<tr>
<td>Kniebel 1988</td>
<td>RCT comparing a combination of hypericum and valerian vs. amitriptyline in depressive patients</td>
</tr>
<tr>
<td>Kugler 1990a</td>
<td>RCT comparing hypericum and bromazepam in depressed patients (included in earlier versions of this review; inclusion has now been limited to commonly recommended standard antidepressants)</td>
</tr>
<tr>
<td>Kugler 1990b</td>
<td>RCT on pharmacodynamics in depressed patients (reason for exclusion: no symptomatic outcomes related to depressive symptoms)</td>
</tr>
<tr>
<td>König 1993</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Lehrl 1993</td>
<td>Diagnosis not described as major depression in publication. An information collected from the sponsor for a previous version of the review that patients met criteria for major depression could not be verified (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Lenoir 1999</td>
<td>RCT comparing three dosages of a hypericum extract in depressive patients</td>
</tr>
<tr>
<td>Li 2005</td>
<td>Not major depression (placebo-controlled RCT in depressive patients undergoing coronary artery bypass grafting)</td>
</tr>
<tr>
<td>Reference</td>
<td>Details</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Maisenbacher 1995</td>
<td>Not all patients suffered from depression (RCT of hypericum for anxiety)</td>
</tr>
<tr>
<td>Martinez 1993</td>
<td>RCT comparing light therapy and a combination of light therapy and hypericum in patients with seasonal affective disorder</td>
</tr>
<tr>
<td>Mo 2004</td>
<td>Not major depression (RCT comparing an hypericum extract and fluoxetine in post-stroke depression)</td>
</tr>
<tr>
<td>Osterheider 1992</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Panijel 1985</td>
<td>Not depression (RCT of a combination of hypericum and valerian for anxiety)</td>
</tr>
<tr>
<td>Quandt 1993</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Reh 1992</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Schlich 1987</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Schmidt 1989</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Schmidt 1993</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Schmidt 1993b</td>
<td>RCT investigating possible interactions of hypericum and alcohol in healthy volunteers</td>
</tr>
<tr>
<td>Schulz 1993</td>
<td>RCT investigating the effects of hypericum on the sleep-EEG in elderly volunteers</td>
</tr>
<tr>
<td>Sindrup 2000</td>
<td>RCT in patients with polyneuropathy (depression not mentioned)</td>
</tr>
<tr>
<td>Sommer 1994</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Spielberger 1985</td>
<td>RCT comparing two hypericum preparations for depression</td>
</tr>
<tr>
<td>Staffeldt 1993</td>
<td>RCT on pharmacokinetics in healthy volunteers</td>
</tr>
<tr>
<td>Steger 1985</td>
<td>RCT of a comination of of hypericum and valerian vs. desipramine for depression</td>
</tr>
<tr>
<td>Volz 2002</td>
<td>RCT hyreicum vs. placebo in patients with somatoform disorders</td>
</tr>
<tr>
<td>Vorbach 1994</td>
<td>Not all patients met criteria for major depression (comparison with maprotiline; included in previous versions of the review)</td>
</tr>
<tr>
<td>Warnecke 1986</td>
<td>Open controlled trial of hypericum vs. diazepam in women with climacteric depression; method of allocation unclear (included in earlier version of this review; now inclusion limited to explicitly randomized, double-blind trials)</td>
</tr>
<tr>
<td>Werth 1989</td>
<td>RCT of hypericum vs. amitriptyline for about two weeks in patients undergoing amputation to prevent reactive depression (included in earlier version of this review)</td>
</tr>
</tbody>
</table>
(Continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wienert 1991</td>
<td>RCT on photosensitivity after application of a combination of hypericum and valerian</td>
</tr>
<tr>
<td>Winkel 2000</td>
<td>Not major depression (placebo-controlled trial included in previous versions of the review)</td>
</tr>
<tr>
<td>Zeller 2000</td>
<td>(Nonrandomized) comparison of different administration schedules of a hypericum extract in patients with mild to moderate depression (not restricted to major depression)</td>
</tr>
</tbody>
</table>
### DATA AND ANALYSES

#### Comparison 1. Hypericum mono-preparations vs. placebo. A. Dichotomous measures

<table>
<thead>
<tr>
<th>Outcome or subgroup title</th>
<th>No. of studies</th>
<th>No. of participants</th>
<th>Statistical method</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Responder - grouped by precision - primary analysis</td>
<td>18</td>
<td>3064</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.48 [1.23, 1.77]</td>
</tr>
<tr>
<td>1.1 Less precise trials</td>
<td>9</td>
<td>1020</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.87 [1.22, 2.87]</td>
</tr>
<tr>
<td>1.2 More precise trials</td>
<td>9</td>
<td>2044</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.28 [1.10, 1.49]</td>
</tr>
<tr>
<td>2 Responder - according to HAMD</td>
<td>16</td>
<td>2706</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.51 [1.22, 1.87]</td>
</tr>
<tr>
<td>2.1 Less precise trials</td>
<td>8</td>
<td>948</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.94 [1.19, 3.18]</td>
</tr>
<tr>
<td>2.2 More precise trials</td>
<td>8</td>
<td>1758</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.28 [1.06, 1.53]</td>
</tr>
<tr>
<td>3 Responder - according to CGI (Clinical Global Impression Index at least “much improved”)</td>
<td>13</td>
<td>2306</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.47 [1.24, 1.74]</td>
</tr>
<tr>
<td>3.1 Less precise trials</td>
<td>7</td>
<td>869</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.74 [1.30, 2.33]</td>
</tr>
<tr>
<td>3.2 More precise trials</td>
<td>6</td>
<td>1437</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.26 [1.06, 1.50]</td>
</tr>
<tr>
<td>4 Responder - grouped by extract</td>
<td>18</td>
<td>981</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.31 [0.92, 1.86]</td>
</tr>
<tr>
<td>4.1 extract LI 160</td>
<td>6</td>
<td>981</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.31 [0.92, 1.86]</td>
</tr>
<tr>
<td>4.2 extract WS 5570</td>
<td>2</td>
<td>699</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.57 [0.96, 2.56]</td>
</tr>
<tr>
<td>4.3 extract WS 5572</td>
<td>2</td>
<td>170</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.47 [1.05, 2.06]</td>
</tr>
<tr>
<td>4.4 extract STW3-VI</td>
<td>2</td>
<td>401</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>3.59 [0.41, 31.56]</td>
</tr>
<tr>
<td>4.5 other extracts</td>
<td>6</td>
<td>813</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.45 [1.08, 1.93]</td>
</tr>
<tr>
<td>5 Responder among studies from German-speaking countries and other studies</td>
<td>18</td>
<td>3064</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.48 [1.23, 1.77]</td>
</tr>
<tr>
<td>5.1 Studies from German-speaking countries</td>
<td>11</td>
<td>1770</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.78 [1.42, 2.25]</td>
</tr>
<tr>
<td>5.2 Studies from other countries</td>
<td>7</td>
<td>1294</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.07 [0.88, 1.31]</td>
</tr>
<tr>
<td>6 Remission (HAMD score &lt; 8 or &lt; 7)</td>
<td>6</td>
<td>1236</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>2.77 [1.80, 4.26]</td>
</tr>
</tbody>
</table>

#### Comparison 2. Hypericum mono-preparations vs. placebo. B. Continuous measures

<table>
<thead>
<tr>
<th>Outcome or subgroup title</th>
<th>No. of studies</th>
<th>No. of participants</th>
<th>Statistical method</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Mean HAMD (Hamilton Rating Scale for Depression) scores after therapy</td>
<td>17</td>
<td>2871</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-3.04 [-4.29, -1.78]</td>
</tr>
</tbody>
</table>
2 Mean HAMD (Hamilton Rating Scale for Depression) scores after 2 to 3 weeks of treatment
Mean Difference (IV, Random, 95% CI) -1.22 [-2.07, -0.37]

3 Mean HAMD (Hamilton Rating Scale for Depression) score after 4 weeks of treatment
Mean Difference (IV, Random, 95% CI) -1.65 [-2.78, -0.52]

4 Mean HAMD (Hamilton Rating Scale for Depression) scores after 6 to 8 weeks of treatment
Mean Difference (IV, Random, 95% CI) -2.97 [-4.31, -1.63]

5 Difference HAMD (Hamilton Rating Scale for Depression) baseline - end of treatment
Mean Difference (IV, Random, 95% CI) -3.03 [-4.67, -1.39]

6 MADRS after treatment
Mean Difference (IV, Random, 95% CI) -3.86 [-7.30, -0.42]

7 Difference MADRS baseline - end of treatment
Mean Difference (IV, Random, 95% CI) -3.01 [-4.88, -1.14]

8 Mean HAMD after treatment in studies from German-speaking countries and other studies
8.1 Studies from German-speaking countries
Mean Difference (IV, Random, 95% CI) -4.29 [-5.61, -2.97]
8.2 Studies from other countries
Mean Difference (IV, Random, 95% CI) -0.77 [-1.74, 0.20]

9 Mean Depression Scale von Zerssen (D-S) after therapy/difference baseline - after therapy
Mean Difference (IV, Random, 95% CI) -3.47 [-5.32, -2.12]

10 Various self-rating scales
10.1 von Zerssen Depression Scale (D-S) after treatment
Mean Difference (IV, Random, 95% CI) -0.47 [-0.64, -0.30]
10.2 von Zerssen Depression Scale (D-S) difference baseline - after treatment
Mean Difference (IV, Random, 95% CI) -0.62 [-1.11, -0.14]
10.3 von Zerssen Adjective Mood Scale
Mean Difference (IV, Random, 95% CI) -0.90 [-2.02, 0.22]
10.4 Beck Depression Inventory
Mean Difference (IV, Random, 95% CI) -0.61 [-0.84, -0.37]
10.5 Beck Depression Inventory difference baseline - after treatment
Mean Difference (IV, Random, 95% CI) -0.28 [-0.56, 0.00]
10.6 Zung Self Rating Depression Scale (SDS) difference baseline - after treatment
Mean Difference (IV, Random, 95% CI) -0.31 [-0.71, 0.09]
10.7 Symptom Checklist (SCL-58) depression score
Mean Difference (IV, Random, 95% CI) -0.37 [-0.73, -0.02]
10.8 von Zerssen Paranoid-Depressivitäts-Skala
Mean Difference (IV, Random, 95% CI) -0.16 [-0.37, 0.04]

11 Various self-rating scales in studies from German-speaking countries and other countries
Mean Difference (IV, Random, 95% CI) -0.47 [-0.64, -0.30]
### Comparison 3. Safety - Hypericum mono-preparations vs. placebo

<table>
<thead>
<tr>
<th>Outcome or subgroup title</th>
<th>No. of studies</th>
<th>No. of participants</th>
<th>Statistical method</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Number of patients discontinuing treatment/dropping out for adverse effects - primary analysis</td>
<td>16</td>
<td>2784</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.92 [0.45, 1.88]</td>
</tr>
<tr>
<td>2 Number of patients dropping out</td>
<td>16</td>
<td>2784</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.87 [0.67, 1.12]</td>
</tr>
<tr>
<td>3 Number of patients reporting adverse effects</td>
<td>14</td>
<td>2496</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.98 [0.78, 1.23]</td>
</tr>
</tbody>
</table>

### Comparison 4. Hypericum mono-preparations vs. standard antidepressants. A. Dichotomous measures

<table>
<thead>
<tr>
<th>Outcome or subgroup title</th>
<th>No. of studies</th>
<th>No. of participants</th>
<th>Statistical method</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Responder (intent to treat) - primary analysis</td>
<td>17</td>
<td>2810</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.01 [0.93, 1.09]</td>
</tr>
<tr>
<td>1.1 vs. older antidepressants</td>
<td>5</td>
<td>1016</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.02 [0.90, 1.15]</td>
</tr>
<tr>
<td>1.2 vs. SSRIs</td>
<td>12</td>
<td>1794</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.00 [0.90, 1.12]</td>
</tr>
<tr>
<td>2 Responder (per protocol)</td>
<td>17</td>
<td>2306</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>0.96 [0.88, 1.05]</td>
</tr>
<tr>
<td>2.1 vs. older antidepressants</td>
<td>5</td>
<td>854</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>0.93 [0.78, 1.11]</td>
</tr>
<tr>
<td>2.2 vs. SSRIs</td>
<td>12</td>
<td>1452</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>0.97 [0.87, 1.08]</td>
</tr>
<tr>
<td>3 Responder according to CGI (Clinical Global Impression Index at least “much improved”)</td>
<td>12</td>
<td>2234</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.01 [0.94, 1.09]</td>
</tr>
<tr>
<td>3.1 vs. older antidepressants</td>
<td>4</td>
<td>692</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>0.97 [0.87, 1.09]</td>
</tr>
<tr>
<td>3.2 vs. newer antidepressants</td>
<td>8</td>
<td>1542</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.03 [0.92, 1.15]</td>
</tr>
<tr>
<td>4 Responder among studies from German-speaking studies and other studies</td>
<td>17</td>
<td>2769</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.00 [0.93, 1.09]</td>
</tr>
<tr>
<td>4.1 Studies from German-speaking countries</td>
<td>9</td>
<td>1952</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.04 [0.96, 1.13]</td>
</tr>
<tr>
<td>4.2 Studies from other countries</td>
<td>8</td>
<td>817</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>0.90 [0.76, 1.06]</td>
</tr>
<tr>
<td>5 Remission (HAMD score &lt; 8)</td>
<td>4</td>
<td>685</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>1.24 [1.02, 1.50]</td>
</tr>
<tr>
<td>5.1 vs. older antidepressants</td>
<td>0</td>
<td>0</td>
<td>Risk Ratio (M-H, Random, 95% CI)</td>
<td>Not estimable</td>
</tr>
</tbody>
</table>
### Comparison 5. Hypericum mono-preparations vs. standard antidepressants. B. Continuous measures

<table>
<thead>
<tr>
<th>Outcome or subgroup title</th>
<th>No. of studies</th>
<th>No. of participants</th>
<th>Statistical method</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Mean HAMD (Hamilton Rating Scale for Depression) after therapy</td>
<td>12</td>
<td>1889</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.39 [-1.23, 0.45]</td>
</tr>
<tr>
<td>1.1 vs. older antidepressants</td>
<td>3</td>
<td>477</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.06 [-1.82, 1.71]</td>
</tr>
<tr>
<td>1.2 vs. SSRIs</td>
<td>9</td>
<td>1412</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.52 [-1.55, 0.51]</td>
</tr>
<tr>
<td>2 Mean HAMD (Hamilton Rating Scale for Depression) scores after 2 or 3 weeks of treatment</td>
<td>9</td>
<td>1529</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.12 [-1.02, 0.78]</td>
</tr>
<tr>
<td>2.1 vs. older antidepressants</td>
<td>3</td>
<td>477</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.05 [-1.31, 1.20]</td>
</tr>
<tr>
<td>2.2 vs. SSRIs</td>
<td>6</td>
<td>1052</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.25 [-1.50, 1.00]</td>
</tr>
<tr>
<td>3 Mean HAMD (Hamilton Rating Scale for Depression) scores after 4 weeks of treatment</td>
<td>9</td>
<td>1367</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.34 [-1.48, 0.80]</td>
</tr>
<tr>
<td>3.1 vs. older antidepressants</td>
<td>3</td>
<td>477</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>0.02 [-1.11, 1.15]</td>
</tr>
<tr>
<td>3.2 vs. SSRIs</td>
<td>6</td>
<td>890</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.69 [-2.44, 1.06]</td>
</tr>
<tr>
<td>4 Mean HAMD (Hamilton Rating Scale for Depression) scores after 6 to 8 weeks of treatment</td>
<td>10</td>
<td>1659</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.34 [-1.24, 0.57]</td>
</tr>
<tr>
<td>4.1 vs. older antidepressants</td>
<td>2</td>
<td>391</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.21 [-2.56, 2.14]</td>
</tr>
<tr>
<td>4.2 vs. SSRIs</td>
<td>8</td>
<td>1268</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.38 [-1.46, 0.69]</td>
</tr>
<tr>
<td>5 Difference HAMD (Hamilton Rating Scale for Depression) baseline - end of treatment</td>
<td>10</td>
<td>1652</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.35 [-1.23, 0.52]</td>
</tr>
<tr>
<td>5.1 vs. older antidepressants</td>
<td>1</td>
<td>210</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-1.20 [-3.29, 0.89]</td>
</tr>
<tr>
<td>5.2 vs. SSRIs</td>
<td>9</td>
<td>1442</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.25 [-1.21, 0.71]</td>
</tr>
<tr>
<td>6 MADRS after treatment</td>
<td>1</td>
<td>108</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.90 [-4.73, 2.93]</td>
</tr>
<tr>
<td>6.1 vs. older antidepressants</td>
<td>0</td>
<td>0</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>6.2 vs. SSRIs</td>
<td>1</td>
<td>108</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.90 [-4.73, 2.93]</td>
</tr>
<tr>
<td>7 Difference MADRS baseline - end of treatment</td>
<td>2</td>
<td>352</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-2.90 [-5.10, -0.70]</td>
</tr>
<tr>
<td>7.1 vs. older antidepressants</td>
<td>0</td>
<td>0</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>7.2 vs. SSRIs</td>
<td>2</td>
<td>352</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-2.90 [-5.10, -0.70]</td>
</tr>
<tr>
<td>8 Mean HAMD after treatment in studies from German-speaking countries and other studies</td>
<td>15</td>
<td>2423</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.39 [-1.23, 0.45]</td>
</tr>
<tr>
<td>8.1 Studies from German-speaking countries</td>
<td>9</td>
<td>1888</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.43 [-1.28, 0.41]</td>
</tr>
<tr>
<td>8.2 Studies from other countries</td>
<td>6</td>
<td>535</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>-0.44 [-2.67, 1.79]</td>
</tr>
<tr>
<td>9 Mean D-S (Depression Scale von Zerssen) scores after therapy</td>
<td>4</td>
<td>360</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>2.66 [0.83, 4.50]</td>
</tr>
<tr>
<td>9.1 vs. older antidepressants</td>
<td>2</td>
<td>272</td>
<td>Mean Difference (IV, Random, 95% CI)</td>
<td>2.81 [0.77, 4.85]</td>
</tr>
</tbody>
</table>
Comparison 6. Safety - Hypericum mono-preparations vs. standard antidepressants

<table>
<thead>
<tr>
<th>Outcome or subgroup title</th>
<th>No. of studies</th>
<th>No. of participants</th>
<th>Statistical method</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Number of patients discontinuing treatment/ dropping out due to adverse/ side effects</td>
<td>16</td>
<td>2785</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.41 [0.29, 0.60]</td>
</tr>
<tr>
<td>1.1 vs. older antidepressants</td>
<td>5</td>
<td>1016</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.24 [0.13, 0.46]</td>
</tr>
<tr>
<td>1.2 vs. SSRIs</td>
<td>11</td>
<td>1769</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.53 [0.34, 0.83]</td>
</tr>
<tr>
<td>2 Number of patients dropping out</td>
<td>16</td>
<td>2785</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.77 [0.62, 0.95]</td>
</tr>
<tr>
<td>2.1 vs. older antidepressants</td>
<td>5</td>
<td>1016</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.67 [0.47, 0.95]</td>
</tr>
<tr>
<td>2.2 vs. SSRIs</td>
<td>11</td>
<td>1769</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.83 [0.63, 1.08]</td>
</tr>
<tr>
<td>3 Number of patients reporting adverse effects</td>
<td>14</td>
<td>2663</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.56 [0.43, 0.74]</td>
</tr>
<tr>
<td>3.1 vs. older antidepressants</td>
<td>5</td>
<td>1016</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.39 [0.30, 0.50]</td>
</tr>
<tr>
<td>3.2 vs. SSRIs</td>
<td>9</td>
<td>1647</td>
<td>Odds Ratio (M-H, Random, 95% CI)</td>
<td>0.70 [0.49, 1.00]</td>
</tr>
</tbody>
</table>
WHAT'S NEW

Last assessed as up-to-date: 7 August 2008.

8 February 2009  Amended  Contact details updated

HISTORY


Review first published: Issue 4, 1998

15 July 2008  New citation required and conclusions have changed

1) The title of the review has been changed from 'St. John's wort for depression' to 'St. John's wort for major depression'. This reflects that inclusion has been now limited to trials in patients suffering from major depression only.

2) The modification of selection criteria resulted in the exclusion of 16 of 37 studies included in the previous version. Eight new trials have been included. The review now covers 29 trials with 18 comparisons of a hypericum extract with a placebo and 17 with a standard antidepressant (six three-armed trials).

3) Compared to the previous version our conclusions are now slightly more favourable, as modest effects over placebo have also been shown in several large trials, and as side effects appear to be less frequent compared to both older antidepressants and selective serotonin reuptake inhibitors.

4) The list of authors has been amended.

23 May 2008  Amended  Converted to new review format.

12 December 2007  New citation required and conclusions have changed  Substantive amendment
CONTRIBUTIONS OF AUTHORS
Study concept and design: Linde, Berner, Kriston
Acquisition of data: Linde, Berner, Kriston
Analysis and interpretation of data: Linde, Berner, Kriston
Drafting and revision of the manuscript: Linde, Berner, Kriston
Statistical expertise: Kriston
Study coordination: Linde

DECLARATIONS OF INTEREST
Michael Berner has received a grant for research on hypericum in the past and has received fees for speaking on conferences from a manufacturer (Schwabe). Klaus Linde once received reimbursement for travel expenses for speaking on a meeting organised by a manufacturer (Schwabe). Levente Kriston once received reimbursement for travel expenses (Schwabe).

SOURCES OF SUPPORT

Internal sources
- Centre for Complementary Medicine Research, Department of Internal Medicine II, Technische Universität München, Germany.
- Department of Psychiatry and Psychotherapy, University of Freiburg, Germany.

External sources
- No sources of support supplied

INDEX TERMS

Medical Subject Headings (MeSH)
*Hypericum; *Phytotherapy; Antidepressive Agents [therapeutic use]; Depressive Disorder [* drug therapy]; Plant Extracts [*therapeutic use]; Randomized Controlled Trials as Topic

MeSH check words
Adult; Humans