NATIONAL DEFENSE RESEARCH INSTITUTE

St. John's Wort for Major Depressive Disorder

A Systematic Review

Alicia Ruelaz Maher, Susanne Hempel, Eric Apaydin, Roberta M. Shanman, Marika Booth, Jeremy N. V. Miles, Melony E. Sorbero

Prepared for the Office of the Secretary of Defense

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2015		**00-00-2015 to 00-00-2015**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
St. John's Wort for Major Depressive Disorder: A Systematic Review	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **RAND Corporation,National Defense Research Institute,1776 Main Street, P.O. Box 2138,Santa Monica,CA,90407-2138**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This systematic review synthesized evidence of St. John???s wort (SJW) for the treatment of major depressive disorder (MDD) (PROSPERO record CRD42015016406). In November 2014, we searched nine electronic databases, as well as bibliographies of existing systematic reviews, to identify randomized controlled trials (RCTs) testing the efficacy and safety of SJW to treat adults with MDD. Two independent reviewers screened publications using predetermined eligibility criteria, abstracted study-level information, and assessed the quality of included studies. Outcomes of interest included changes in depressive symptomatology, quality of life, and adverse effects. Efficacy meta-analyses used the Hartung-Knapp-Sidik-Jonkman method for random-effects models. Quality of evidence was assessed using the GRADE approach. In total, 35 studies met inclusion criteria. There is moderate evidence, due to unexplained heterogeneity between studies, that depression improvement based on the number of treatment responders (RR 0.65; CI 0.51, 0.84; I2 79%; 18 RCTs; n=2,922) and depression scale scores (SMD 0.49; CI 0.23, 0.74; 16 RCTs; I2 89%, n=2,888) favors SJW over placebo. There is low quality evidence of no statistically significant difference in the number of patients in remission (RR 0.60; CI 0.22, 1.66; 9 RCTs; I2 94%). The existing evidence is based on RCTs testing SJW as monotherapy; there is a lack of evidence for SJW given as adjunct therapy to standard antidepressant treatment. We found no systematic difference between SJW extracts, but head-tohead trials are missing; LI of 160 (0.3% hypericin, 1???4% hyperforin) was the extract with the greatest number of RCTs. The existing research is primarily based on combined mild and moderate depression patient samples, and there is a lack of research in severe depression. Only two RCTs assessed quality of life. There is moderate evidence that SJW participants are not more likely than placebo participants to experience adverse events generally (OR 0.83; CI 0.62 1.13; 13 RCTs), but SJW was associated with more neurologic and organ system (e.g., renal) adverse events, and assessments were limited and inadequate for rare events. Comparing SJW with antidepressant medication showed moderate evidence for patients on antidepressants experiencing more adverse events (OR 0.67; CI 0.56, 0.81; 11 RCTs) and low evidence that SJW is associated with fewer specific adverse events, including gastrointestinal and neurologic adverse events. There were no systematic differences in responders (RR 0.99; CI 0.88, 1.11; 17 RCTs, I2 53%; moderate evidence), depression scores (SMD 0.03; CI ???0.15, 0.21; 14 RCTs; I2 74%; moderate evidence), or remission (RR 0.86; CI 0.61, 1.20; 7 RCTs; I2 29%; low evidence). SJW monotherapy for mild and moderate depression was superior to placebo in improving symptoms and not

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**144**	

For more information on this publication, visit www.rand.org/t/RR1048

Published by the RAND Corporation, Santa Monica, Calif.

Preface

The Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury is interested in determining the efficacy and comparative effectiveness of integrative medicine approaches for psychological health conditions. This document is a systematic review of the effectiveness of St. John's wort for major depressive disorder (MDD), conducted during year two of a two-year project on integrative medicine approaches for psychological health conditions. The review will be of interest to military health policymakers and practitioners, civilian health care providers, and policymakers, payers, and patients.

A version of this report was provided to the committee for review in April 2015; we reproduce that version here, with minor editorial updates. None of the authors has any conflict of interest to declare.

This research was sponsored by the Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury and conducted within the Forces and Resources Policy Center of the RAND National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community. For more information on the RAND Forces and Resources Policy Center, see http://www.rand.org/nsrd/ndri/centers/frp.html or contact the director (contact information is provided on the web page).

Abstract

This systematic review synthesized evidence of St. John's wort (SJW) for the treatment of major depressive disorder (MDD) (PROSPERO record CRD42015016406).

In November 2014, we searched nine electronic databases, as well as bibliographies of existing systematic reviews, to identify randomized controlled trials (RCTs) testing the efficacy and safety of SJW to treat adults with MDD. Two independent reviewers screened publications using predetermined eligibility criteria, abstracted study-level information, and assessed the quality of included studies. Outcomes of interest included changes in depressive symptomatology, quality of life, and adverse effects. Efficacy meta-analyses used the Hartung-Knapp-Sidik-Jonkman method for random-effects models. Quality of evidence was assessed using the GRADE approach.

In total, 35 studies met inclusion criteria. There is moderate evidence, due to unexplained heterogeneity between studies, that depression improvement based on the number of treatment responders (RR 0.65; CI 0.51, 0.84; I^2 79%; 18 RCTs; n=2,922) and depression scale scores (SMD 0.49; CI 0.23, 0.74; 16 RCTs; I^2 89%, n=2,888) favors SJW over placebo. There is low quality evidence of no statistically significant difference in the number of patients in remission (RR 0.60; CI 0.22, 1.66; 9 RCTs; I^2 94%). The existing evidence is based on RCTs testing SJW as monotherapy; there is a lack of evidence for SJW given as adjunct therapy to standard antidepressant treatment. We found no systematic difference between SJW extracts, but head-to-head trials are missing; LI of 160 (0.3% hypericin, 1–4% hyperforin) was the extract with the greatest number of RCTs. The existing research is primarily based on combined mild and moderate depression patient samples, and there is a lack of research in severe depression. Only two RCTs assessed quality of life. There is moderate evidence that SJW participants are not more likely than placebo participants to experience adverse events generally (OR 0.83; CI 0.62, 1.13; 13 RCTs), but SJW was associated with more neurologic and organ system (e.g., renal) adverse events, and assessments were limited and inadequate for rare events. Comparing SJW with antidepressant medication showed moderate evidence for patients on antidepressants experiencing more adverse events (OR 0.67; CI 0.56, 0.81; 11 RCTs) and low evidence that SJW is associated with fewer specific adverse events, including gastrointestinal and neurologic adverse events. There were no systematic differences in responders (RR 0.99; CI 0.88, 1.11; 17 RCTs, I^2 53%; moderate evidence), depression scores (SMD 0.03; CI −0.15, 0.21; 14 RCTs; I^2 74%; moderate evidence), or remission (RR 0.86; CI 0.61, 1.20; 7 RCTs; I^2 29%; low evidence).

SJW monotherapy for mild and moderate depression was superior to placebo in improving symptoms and not significantly different from antidepressant medication, but there was evidence of heterogeneity. SJW adverse events reported in included RCTs were comparable to placebo

and fewer compared with antidepressant medication; however, adverse event assessments were limited, and thus we have limited confidence in this conclusion.

Table of Contents

Figures

Tables

Summary

Introduction

Worldwide, depressive disorders are one of the largest sources of disease burden. Depression is commonly treated with prescription medications. However, these can cause side effects, and many patients turn to alternative treatments, such as St. John's wort (SJW). Previous systematic reviews have shown the efficacy of SJW in mild to moderate depression, but additional studies have been completed since that time. This systematic review aims to synthesize evidence from trials of SJW to provide estimates of its effectiveness in treating major depressive disorder (MDD) (PROSPERO record CRD42015016406).

This review was guided by the following key questions (KQs):

- *KQ 1:* What are the efficacy and safety of St. John's wort (SJW), as an adjunctive or monotherapy, for depressive symptoms and quality of life in adults with MDD compared with placebo or active comparator?
  - *KQ 1a:* Is SJW more effective as monotherapy than as an adjunctive therapy?
  - *KQ 1b:* Is there a difference in efficacy, depending on the amount and type of extract of SJW used?
  - *KQ 1c:* Is there a difference in efficacy, depending on the type of MDD (i.e., mild, moderate, or severe)?
  - *KQ 1d:* Are adverse events associated with SJW comparable to standard antidepressant treatment?
  - *KQ 1e:* Is the efficacy of SJW comparable to standard antidepressant treatment?

Methods

To answer our key questions, we conducted a systematic search of electronic databases (PubMed, CINAHL, PsycINFO, CENTRAL, Embase, AMED, MANTIS, Web of Science, and ICTRP) without language restriction to November 2014, as well as bibliographies of existing systematic reviews and included studies, to identify reports of randomized controlled trials (RCTs) testing the efficacy and safety of SJW—used adjunctively or as monotherapy—to treat adults with MDD.

Two independent reviewers screened the identified literature using predetermined eligibility criteria, abstracted prespecified study-level information, and assessed the quality of included studies. Outcomes of interest included changes in depressive symptomatology, quality of life, and adverse effects.

Meta-analyses for efficacy outcomes and the number of patients with adverse events were conducted using the Hartung-Knapp-Sidik-Jonkman method for random-effects models to estimate the relative risk (RR) and standardized mean differences (SMDs), together with the 95-percent confidence interval (CI). For specific adverse events, many of which are very rare, we used exact conditional methods to estimate odd ratios (ORs). The quality of evidence was assessed using the Grades of Recommendation, Assessment, Development, and Evaluation (or GRADE) approach.

Results

In total, 35 studies met inclusion criteria. All studies reported on the efficacy of SJW, and 34 addressed safety. Risk of bias in included studies varied: Ten studies obtained a "good" quality rating, 14 studies were rated "fair," and 11 were rated "poor" quality.

Key Question 1

We found moderate evidence (due to unexplained heterogeneity between studies) that, compared with placebo, SJW is associated with improvement in depression symptoms. SJW groups reported significantly more treatment responders, usually defined by study authors as a 50-percent reduction in Hamilton Rating Scale for Depression scores (RR 0.65; CI 0.51, 0.84; I^2 79%; 18 RCTs; n=2,922), and participants receiving SJW had significantly lower depression scale scores (SMD 0.49; CI 0.23, 0.74; 16 RCTs; I^2 89%, n=2,888) than participants receiving a placebo. Sensitivity analyses showed very similar results when excluding poor quality studies. There is low quality evidence of no statistically significant differences in the number of patients in remission (RR 0.60; CI 0.22 to 1.66; 10 RCTs; I^2 94%).

Only two studies assessed quality of life and compared effects of SJW with placebo or with standard antidepressant medication.

Most (34 of 35) of the included studies addressed the safety of SJW, but rigor of assessment varied greatly. In the included RCTs, there was moderate evidence that SJW is not more likely to cause adverse events than placebo, overall (OR 0.83; CI 0.62, 1.13; 13 RCTs). However, specific adverse events, such as neurologic/nervous system and organ system (e.g., eye, ear, liver, renal, reproductive) adverse events, were more likely in those taking SJW. Furthermore, the adverse events assessments were limited and inadequate for rare adverse events.

Key Question 1a

We found only one study examining the use of SJW used adjunctive to standard antidepressant treatment (medication or psychotherapy). Therefore, the review is unable to assess whether SJW is more effective as monotherapy than as an adjunctive therapy. The existing evidence for SJW is based on monotherapy research.

Key Question 1b

We found only one study that compared two different standardized extracts and three studies that compared different dosages, none of which found statistically significant differences. Several studies did not specify the extract of SJW used. Of those that did, the most common extract was LI 160 (0.3% hypericin, 1–4% hyperforin). A comparison across studies did not indicate systematic differences in outcomes depending on the extract used (outcome responders p=0.347; depression scale scores p=0.127; remission p=0.371).

Key Question 1c

Analyses did not suggest that the effectiveness or safety of SJW varies by depression severity, but the existing research is primarily based on combined mild and moderate depression patient samples and there is a lack of research studies in severe depression.

The review did not find sufficient evidence to estimate the treatment effect of SJW compared with placebo for mild depression alone or severe depression alone. Three studies provided results for patients with moderate depression compared with placebo and found statistically significant effects in the number of responders and continuous depression outcome in the individual studies, but confidence intervals in the pooled analyses did not suggest a statistically significant difference compared with placebo. The treatment effects in the largest subgroup (combined mild and moderate depression) were similar to the main analysis that included all studies, and a meta-regression did not show statistically significant effects of an association between the depression severity and the size of the treatment effect of SJW compared with placebo (outcome responders p=0.798; depression scale scores p=0.365; remission p=0.159).

Key Question 1d

In the included monotherapy RCTs comparing SJW with standard antidepressant medications, there was moderate evidence that those patients taking antidepressants experienced more adverse events overall (OR 0.67; CI 0.56, 0.81; 11 RCTs). Compared with such antidepressants as selective serotonin reuptake inhibitors (SSRIs), there was low quality evidence showing that SJW is associated with fewer specific adverse events, including gastrointestinal (OR 0.43; CI 0.34, 0.55; 15 RCTs) and neurologic (OR 0.29; CI 0.24 to 0.36; 15 RCTs) adverse events. We identified only one study reporting a comparison with psychotherapy. The rigor of adverse event assessments and reporting varied greatly, comparative analyses were potentially limited due to the lack of statistical power to show differences in individual rare events, and RCTs addressed only a limited range of potential adverse events.

Key Question 1e

We found no systematic differences in treatment responders (RR 0.99; CI 0.88, 1.11; 17 RCTs, I^2 53%; moderate evidence), depression scale scores (SMD 0.03; CI −0.15, 0.21; 14

RCTs; I^2 74%; moderate evidence), or patients in remission (RR 0.86; CI 0.61, 1.20; 7 RCTs; I^2 29%; low evidence) comparing SJW and antidepressant medications used to treat adults with mild or moderate depression. The effects for the outcome responders and depression scale scores remained stable when limiting analysis to RCTs reporting a power calculation and with sufficient statistical power to identify an effect. However, the quality of these identified studies was limited. Studies reporting on remission were also limited in study quality, and the statistical power to detect differences between interventions was unclear. Only one study compared SJW and psychotherapy. There is a lack of data on quality of life. The included studies showed the efficacy of SJW as comparable to antidepressant medication, with SJW being neither inferior nor superior for the treatment of mild or moderate depression.

Conclusions

The review showed SJW given as monotherapy for mild and moderate depression is superior to placebo in improving symptoms and not significantly different from antidepressant medication; however, there was evidence of substantial heterogeneity between studies. SJW adverse events reported in included RCTs were comparable to placebo groups, and there were fewer compared with antidepressant medication; however, adverse event assessments were limited and inadequate for rare events, and thus we have limited confidence in this conclusion.

Acknowledgments

This research is sponsored by the Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury (DCoE). We gratefully acknowledge Kristie Gore for her support and guidance throughout the project, and Christian Lopez, Tanja Perry, Patty Smith, Aneesa Motala, and Ryan Kandrack for research assistance. We also would like to thank our project officers and points of contact at DCoE—Chris Crowe, Marina Khusid, and Michael Freed—for their support of our work. In addition, we thank Paul Shekelle and Klaus Linde for reviewing the report and their helpful suggestions. Any errors of fact or interpretation in this report remain the responsibility of the authors.

Abbreviations

AE	adverse event
AMED	Allied and Complementary Health Database
BDI	Beck Depression Inventory
CENTRAL	Cochrane Central Register of Controlled Trials
CGI	Clinical Global Impression scale
CI	confidence interval
DSM	*Diagnostic and Statistical Manual of Mental Disorders*
GRADE	Grades of Recommendation, Assessment, Development and Evaluation
HAMD	Hamilton Rating Scale for Depression
HDTSG	Hypericum Depression Trial Study Group
ICD	International Classification of Diseases
ICTRP	International Clinical Trials Registry Platform
ITT	intention to treat
KQ	key question
MADRS	Montgomery-Åsberg Depression Rating Scale
MANTIS	Manual, Alternative, and Natural Therapy Index System
MAO	monoamine oxidase
MDD	major depressive disorder
OR	odds ratio
RCT	randomized controlled trial
RR	relative risk
SD	standard deviation
SE	standard error
SJW	St. John's wort
SMD	standardized mean difference
SSRI	selective serotonin reuptake inhibitor
URTI	upper respiratory tract infection

Chapter One: Introduction

Background and Objective

Major depressive disorder (MDD) is a common, serious mental health condition (Ustun et al., 2004). Globally, depressive disorders are the leading cause of disability and a major contributor to the global burden of disease. Worldwide, more than 350 million people suffer from depression, and this number is on the rise (World Health Organization, 2012). The condition affects approximately 15 million individuals in the United States, with a lifetime prevalence of 8 to 12 percent in men and 20 to 26 percent in women, yet the condition remains underdiagnosed and undertreated, particularly among active duty military personnel and veterans (U.S. Department of Veterans Affairs and U.S. Department of Defense, 2009).

Pharmacotherapy and psychotherapy are available and have been shown to be effective to treat depressive disorders, such as MDD. However, stigma, cost, and lack of availability of mental health treatment; side effects of medication; and other factors cause many individuals to not seek standard treatments. An estimated 37 to 72 percent of military personnel use complementary and alternative medicine (Institute of Medicine, 2012). About one-third of them use complementary and alternative medicine for psychological conditions (McPherson and Schwenka, 2004).

For centuries, extracts of the herb St. John's wort (botanical name *Hypericum perforatum L.*, SJW) have been used to treat various conditions, including depressive disorders. A Cochrane Review of SJW for depression documented available research studies published to 2008 and found a beneficial effect compared with both placebo and other antidepressant therapies (Linde, Berner, and Kriston, 2008). Overall, SJW has been considered safe, but side effects have been noted, including photosensitivity, elevated thyroid stimulating hormones, hypertensive crisis, and induction of mania (Knuppel and Linde, 2004). In addition, preparations of SJW vary in the amounts of active compounds they contain, which may make it difficult to compare across studies (Liu, Ang, et al., 2000).

Existing clinical guidelines vary in their recommendations to include SJW as a treatment option for treating depressive disorders (Linde et. al, 2015). The (U.S. Department of Veterans Affairs and U.S. Department of Defense (2009) *Clinical Practice Guideline for Management of Major Depressive Disorder* recommends that SJW may be used by patients with mild MDD who have a strong preference for herbal treatments. However, the guideline also states that it is not recommended for patients with moderate to severe major depression. Furthermore, SJW should not be used by patients taking medication whose clearance is substantially dependent on the CYP3A4 isoenzyme, and SJW is contraindicated in pregnancy. Finally, patients should be

informed of potential drug-drug interactions and advised to inform all prescribing clinicians that they are using this herbal treatment.

In recent years, more research on SJW has been published investigating not only effectiveness and safety but also its comparative effectiveness and safety (e.g., compared with standard antidepressants). This review aims to synthesize studies identified in the 2008 Cochrane review (Linde, Berner, and Kriston, 2008) and current research in a comprehensive systematic review in order to provide reliable estimates of the effectiveness and safety of SJW for MDD.

Key Questions

We conducted a systematic review to identify randomized controlled trials (RCTs) testing the efficacy and safety of SJW to treat individuals with MDD (PROSPERO record CRD42015016406). Specifically, this systematic review aimed to answer the following key questions (KQs):

- *KQ 1:* What are the efficacy and safety of St. John's wort (SJW), as an adjunctive or monotherapy, for depressive symptoms and quality of life in adults with MDD compared with placebo or active comparator?
  - *KQ 1a:* Is SJW more effective as monotherapy than as an adjunctive therapy?
  - *KQ 1b:* Is there a difference in efficacy, depending on the amount and type of extract of SJW used?
  - *KQ 1c:* Is there a difference in efficacy, depending on the type of MDD (i.e., mild, moderate, or severe)?
  - *KQ 1d:* Are adverse events associated with SJW comparable to standard antidepressant treatment?
  - *KQ 1e:* Is the efficacy of SJW comparable to standard antidepressant treatment?

Chapter Two: Methods

Search Strategy

We searched the electronic databases PubMed, CINAHL (Cumulative Index to Nursing and Allied Health Literature), PsycINFO, CENTRAL (Cochrane Central Register of Controlled Trials), Embase, AMED (Allied and Complementary Health Database), MANTIS (Manual, Alternative, and Natural Therapy Index System), Web of Science, and ICTRP (International Clinical Trials Registry Platform) without language restriction from 2007 to November 2014 to identify reports of RCTs. The choice of the initiation point for the searches is based on the release of a Cochrane review by Linde, Berner, and Kriston (2008) covering trials on SJW for MDD to July 2007. The review should have captured all pertinent trials at that time, and we used it to identify relevant studies published prior to 2007 by screening publications included and excluded in the Cochrane review. Our search was not limited to peer-reviewed literature; we included grey literature, such as conference abstracts. In addition, we screened other existing systematic reviews on the topic to ensure that we identified all studies meeting our inclusion criteria (see below).

The search strings for each database were developed by the chief reference librarian for RAND's Knowledge Services, informed by the Cochrane review on SJW (Linde, Berner, and Kriston, 2008). The draft search strategy is shown in Appendix A.

Eligibility Criteria

The inclusion and exclusion criteria for this review were developed using the framework of participants, interventions, comparators, outcomes, timing, settings, and study design, or PICOTSS.

- *Participants*: Studies were limited to adults, male and female, 18 years of age and over, with a diagnosis of MDD. In studies not referring to a clinical diagnosis based on *Diagnostic and Statistical Manual of Mental Disorders* (DSM) or International Classification of Diseases (ICD) criteria, we applied a specified threshold on validated depression scales (see Appendix B). Studies that enrolled individuals with other comorbid conditions, such as traumatic brain injury, were included. Studies in participants in postnatal depression were included if the criteria were in accordance with DSM-V criteria for MDD (i.e., peripartum onset or four weeks following delivery). Studies in individuals with diagnoses of dysthymia, bipolar disorder, or schizophrenia, alone or in combination with major depression, were excluded in accordance with DSM-V criteria. Studies evaluating multiple psychiatric conditions were included if the data for patients with MDD were presented separately.

- *Interventions*: Studies that administered a supplement that contained a known amount of SJW, and the amount and type of active compounds contained in the SJW supplement was specified (i.e., naphthodianthrones, hypericin, pseudohypericin, flavonoids, phloroglucinols, hyperforin and adhyperforin), were included. SJW could be evaluated alone, or in conjunction with pharmacologic and/or psychotherapy.
- *Comparator*: Studies comparing SJW with placebo or with active comparators, or against another amount or extract of SJW, were included.
- *Outcomes*: Studies that reported Hamilton Rating Scale for Depression (HAMD) scores or other validated depression scale scores were included. Studies that reported other changes in depressive symptoms were included (e.g., suicidal ideation or risk for suicide). Studies that reported rates of depression relapse were included. Studies that reported quality-of-life assessment scores, such as the SF-36, were included if the studies also assessed changes in depression. Studies that reported adverse events in adults taking SJW for MDD were included if adverse events were reported by study arm. Studies that reported on biomarkers alone without reporting efficacy for depression outcomes were not included. Studies of provider outcomes, acceptance, prevalence, use, costs, study design features, and intervention features not reporting patient health outcomes were excluded.
- *Timing*: Only studies with a treatment duration of four weeks or longer were included.
- *Setting*: Studies were not limited by setting (e.g., country, physical location of treatment).
- *Study design*: Included studies were limited to RCTs.

Inclusion Screening

Two independent reviewers (the project lead, who is an experienced systematic reviewer, and a RAND doctoral candidate and assistant policy analyst with experience in systematic reviews) screened titles and abstracts of retrieved citations. An initial session piloting the screening form occurred prior to these reviews to ensure similar interpretation of the inclusion and exclusion criteria. Citations judged as potentially eligible by one or both reviewers were obtained as full text. The full-text publications were then screened against the specified inclusion criteria by the two independent reviewers; any disagreements were resolved through discussion within the review author team.

Studies on the same participants were counted as one study regardless of the number of publications the results were presented in. All publications were considered and used for data extraction.

Data Extraction

The two aforementioned reviewers abstracted study-level data in electronic databases. Categorical data were abstracted independently by both reviewers; free text information and adverse events were abstracted by one reviewer and checked by the review lead. Data collection forms were designed by the review team. The reviewers pilot-tested the data collection forms

prior to data extraction to ensure agreement of interpretation. Effectiveness outcome data were abstracted and checked by a biostatistician of the RAND Evidence-based Practice Center (EPC).

The following information was abstracted from each study:

- *Participants*: number; diagnostic criteria, baseline HAMD (or other measure of depression severity), and depression history; comorbidities (including traumatic brain injury); mean age and age range; gender; and study inclusion and exclusion criteria
- *Interventions*: details including amount and type of active compounds contained in the SJW supplement and how the concentrations of active ingredient(s) have been assessed); dosage; co-intervention(s), if any; and washout period, if any
- *Comparators*: type of comparator
- *Outcomes assessed*: assessment measures and primary endpoint, method of data expression (e.g., standardized mean difference, proportion of patients reporting improvement above a minimum clinically important difference), and corresponding results (effect estimate, precision)
- *Timing*: time-points of outcome assessment, duration of intervention, and follow-up assessment
- *Setting*: geographic region, number of sites
- *Study design*: aim of study, inclusion and exclusion criteria, sample size and reported power calculations, and items relevant to risk of bias and quality ratings.

We relied on published data, which could include conference abstracts and dissertations; no inquiries were made to authors or sponsors. Outcome data were based on intention-to-treat (ITT) analyses reported in the included studies. In the absence of ITT data, we used the number randomized as the denominator; in the absence of the number randomized, we used the number of participants at follow-up. All studies were analyzed using the latest reported follow-up; however, follow-up studies reporting follow-up only from treatment responders were not considered. When multiple depression measures were available, we used HAMD scores to assess treatment effects on depression symptoms. We used the authors' definition of responder, usually reflecting a 50-percent decrease in HAMD. We used the authors' definition of remission, usually reflecting a HAMD score of less than seven or eight. We computed standardized mean differences (SMDs) for studies reporting continuous outcomes, relative risks (RRs) for treatment effect estimates, and odds ratios (ORs) for rare adverse events, together with the 95-percent confidence interval (CI).

In accordance with data-sharing conventions, the raw data can be obtained from the authors.

Risk of Bias

The two reviewers assessed the risk of bias of included studies using the Cochrane Risk of Bias tool (Higgins et al., 2011) and quality criteria used by the U.S. Preventative Services Task Force (U.S. Preventive Services Task Force, 2008). Specifically, the reviewers assessed risks of bias related to the following domains: random sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and providers (performance bias), blinding

of outcome assessors (detection bias), completeness of reporting outcome data (attrition bias), and selective outcome reporting (reporting bias). In addition, we assessed whether both treatment arms received treatment as usual with the treatment group receiving SJW and the control group receiving no additional treatment ("add-on trial"). Furthermore, appropriate washout periods or exclusion of individuals taking personal supplements were assessed.

Other biases related to the U.S. Preventative Services Task Force (USPSTF)'s criteria for internal validity of included studies were also assessed, namely those related to equal distribution amongst groups of potential confounders at baseline; cross-overs or contamination between groups; equal, reliable, and valid outcome measurement; clear definitions of interventions; and ITT analysis. These criteria were used to rate the quality of individual included studies using the following guidelines (Lewin Group and ECRI Institute, 2014; U.S. Preventive Services Task Force, 2008):

- *Good*: Comparable groups are initially assembled and maintained throughout the study with at least 80-percent follow-up; reliable, valid measurement is used and applied equally to all groups; interventions are clearly described; all important outcomes are considered; appropriate attention is given to confounders in analysis; and ITT analysis is used.
- *Fair*: One or more of the following issues is found in the study: some though not major differences between groups exist at follow-up; measurement instruments are acceptable but not ideal, though are generally applied equally; some but not all important outcomes are considered; some but not all potential confounders are accounted for in analyses. ITT analysis must be done.
- *Poor*: One or more of the following "fatal flaws" is found in the study: initially assembled groups are not comparable or maintained throughout the study; unreliable or invalid measurements are used or applied unequally across groups; key confounders are given little to no attention in analyses; ITT analysis is not used.

Data synthesis

The primary aim of this systematic review was to determine what effects SJW has on depressive symptoms, quality of life, and adverse events in adults with MDD compared with placebo and active comparators.

When sufficient data were available and clinical heterogeneity was acceptable, we conducted meta-analyses to pool results across included studies for the outcomes of interest. For all efficacy outcomes and the number of patients with adverse events, we used the Hartung-Knapp-Sidik-Jonkman method for our random-effects meta-analysis (Hartung, 1999; Hartung and Knapp, 2001; Sidik and Jonkman, 2006). This approach may be preferred when the number of studies pooled is small and when there is evidence of heterogeneity (IntHout, Ioannidis, and Borm, 2014). It has been shown that the error rates are more robust than the previously used DerSimonian and Laird method (Sánchez-Meca and Marín-Martínez, 2008). For specific adverse

events, many of which are very rare, we used exact conditional methods to estimate ORs and CIs.

Throughout the review, we differentiated effectiveness and comparative effectiveness analyses. Placebo trials were used to estimate the treatment effect of SJW by demonstrating effects that go beyond placebo effects. A further key aim of the review was to determine the comparative effectiveness of SJW compared with standard antidepressant treatment (both psychotherapy or antidepressant medication). Comparative effectiveness results and equivalence assessments of the efficacy and safety took the consistency of effects across individual studies and the statistical power to detect a statistically significant difference between treatment groups into account.

We conducted subgroup analyses to provide indirect evidence based on the identified literature to answer individual review questions, in particular in the absence of head-to-head trials addressing the research questions. Planned subgroup analyses addressed SJW used as monotherapy versus adjunctive therapy, subgroup analyses for different extracts tested in more than one study, and subgroup for different levels of depression severity (i.e., mild, moderate, and severe depression). We conducted meta-regressions to identify effect modifiers and to identify sources of heterogeneity in study results. We conducted sensitivity analyses to test the robustness of main results (e.g., to test effects in studies with sufficient power to detect effect differences between study arms, for analyses with clear outliers, or excluding poor quality studies) (Greenland and Longnecker, 1992; Orsini et al., 2012; Hamling et al., 2008; Higgins et al., 2011). Publication bias was assessed with the Begg and Egger tests; in the case of indications for bias, treatment estimates were estimated using the trim-and-fill method.

Quality of Evidence

The quality of evidence was assessed for major outcomes using the Grades of Recommendation, Assessment, Development, and Evaluation (or GRADE) approach (Balshem et al., 2011). Namely, the body of evidence was assessed based on the following dimensions: study limitations (low, medium, or high), consistency (consistent, inconsistent, or unknown), directness (direct or indirect), and precision (precise or imprecise) (Egger et al., 1997). The quality of the body of evidence was downgraded in the following instances: results were primarily based on studies with substantial limitations; results were inconsistent across individual studies, in the presence of substantial heterogeneity in pooled analyses, and the result was only based on a single study without replication in an independent research study; conclusions were based on indirect evidence (e.g., effects based on subgroup analyses or meta-regressions in the absence of head-to-head comparisons); and pooled results were imprecise estimates of the treatment effect with wide confidence intervals spanning effect sizes with different clinical conclusions.

The quality of evidence was graded on a four-item scale:

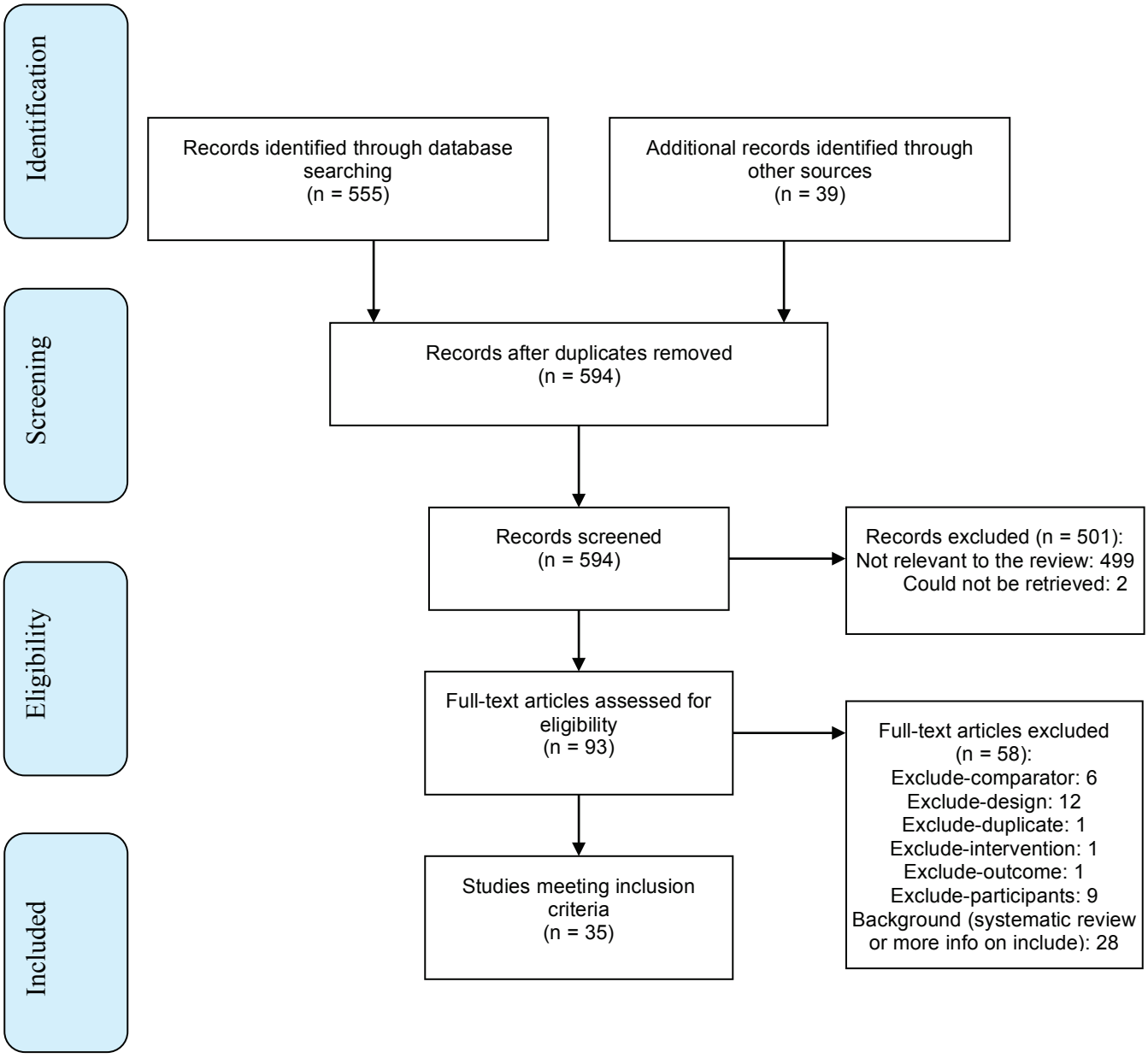
- *High* indicates that the review authors are very confident that the effect estimate lies close to the true effect for a given outcome, as the body of evidence has few or no deficiencies. As such, the reviewers believe the findings are stable. That is, further research is very unlikely to change confidence in the effect estimate.
- *Moderate* indicates that the review authors are moderately confident that the effect estimate lies close to the true effect for a given outcome, as the body of evidence has some deficiencies. As such, the reviewers believe that the findings are likely to be stable, but further research may change confidence in the effect estimate and may even change the estimate.
- *Low* indicates that the review authors have limited confidence that the effect estimate lies close to the true effect for a given outcome, as the body of evidence has major or numerous (or both) deficiencies. As such, the reviewers believe that additional evidence is needed before concluding either that the findings are stable or that the effect estimate lies close to the true effect.
- *Very low* indicates that the review authors have very little confidence that the effect estimate lies close to the true effect for a given outcome, as the body of evidence has very major deficiencies. As such, the true effect is likely to be substantially different from the estimated effect; thus, any estimate of effect is very uncertain.

Chapter Three: Results

Results of the Search

We identified 594 citations through the electronic database search and reference mining of included studies and previous systematic reviews related to SJW (see Figure 3.1).

Figure 3.1. Flow Diagram



Full texts were obtained for 93 citations identified as potentially eligible by the two independent reviewers. In total, 58 articles were excluded at the full-text stage because they did not meet eligibility criteria. We could not obtain two studies to assess them for eligibility. A list of excluded full-text publications is shown in Appendix C. Thirty-five RCTs met inclusion criteria, and details of these studies are available in Appendix D.

Table 3.1. Evidence Base for Key Questions

Key Question		Number of RCTs
KQ 1	What are the efficacy and safety of SJW, as an adjunctive or monotherapy, for depressive symptoms and quality of life in adults with MDD compared with placebo, active comparator, or no treatment?	35 RCTs (19 placebo comparator) (21 active comparator)
KQ 1a	Is SJW more effective as monotherapy than as an adjunctive therapy?	34 monotherapy 1 explicit adjunctive therapy
KQ 1b	Is there a difference in efficacy, depending on which extract of SJW is used?	4 head-to-head trials 8 RCTs testing LI 160 4 RCTs testing WS 5570 3 RCTs testing Ze 117 20 RCTs testing other extracts or not specified
KQ 1c	Is there a difference in efficacy, depending on the type of MDD (i.e., mild, moderate, severe)?	5 moderate 1 severe 20 mild and moderate 2 moderate and severe 8 mixed or not specified
KQ 1d	Are adverse events associated with SJW comparable to standard antidepressant treatment?	19 RCTs
KQ 1e	Is the efficacy of SJW comparable to standard antidepressant treatment?	19 antidepressant comparator

All included studies provided data on the efficacy of SJW, and 34 RCTs addressed the presence or absence of adverse events.

For KQ 1a on whether SJW is more effective as monotherapy than as an adjunctive therapy, we identified only one RCT that utilized SJW systematically as adjunctive therapy (Pakseresht et al., 2012) while the rest of the RCTs studied SJW as monotherapy.

For KQ 1b on whether there is a difference in efficacy, depending on the amount and type of extract of SJW used, we identified four head-to-head trials comparing the effectiveness of different extracts or different amounts or dosing schedules of SJW extract. The most commonly studied extract across all included studies was LI 160 (0.3% hypericin, 1–4% hyperforin), followed by WS 5570 (3–6% hyperforin, 0.1–0.3% hypericin), and Ze 117 (0.2% hyperforin).

Relevant to KQ 1c regarding whether there is a difference in efficacy, depending on the type of MDD (i.e., mild, moderate, severe), is that the majority of studies included samples of participants with mild and/or moderate depression. No study was limited to mild depression only, and only one study tested SJW in severe depression.

For KQ 1d on whether adverse events associated with SJW are comparable to standard antidepressant treatment, we identified 19 RCTs comparing SJW with antidepressant treatment, listing adverse events reported in each treatment arm (Behnke et al., 2002; Bjerkenstedt et al., 2005; Brenner et al., 2000; Fava et al., 2005; Gastpar, Singer, and Zeller, 2005; Gastpar, Singer, and Zeller, 2006; Harrer, Hübner, and Podzuweit, 1994; Harrer et al., 1999; Hypericum Depression Trial Study Group (HDTSG), 2002; Liu et al., 2010; Moreno et al., 2005; Pakseresht et al., 2012; Philipp, Kohnen, and Hiller, 1999; Schrader, 2000; Szegedi et al., 2005; van Gurp et al., 2002; Vorbach, Arnoldt, and Hubner, 1997; Wheatley, 1997; Woelk, 2000).

For KQ 1e on whether the efficacy of SJW is comparable to standard antidepressant treatment, we found 19 RCTs providing data on treatment with SJW versus treatment with standard antidepressants (Behnke et al., 2002; Bjerkenstedt et al., 2005; Brenner et al., 2000; Fava et al., 2005; Gastpar, Singer, and Zeller, 2005; Gastpar, Singer, and Zeller, 2006; HDTSG, 2002; Harrer, Hübner, and Podzuweit, 1994; Harrer et al., 1999; Liu et al., 2010; Moreno et al., 2005; Pakseresht et al., 2012; Philipp, Kohnen, and Hiller, 1999; Schrader, 2000; Szegedi et al., 2005; van Gurp et al., 2002; Vorbach, Arnoldt, and Hubner, 1997; Wheatley, 1997; Woelk, 2000).

Description of Included Studies

Design

All RCTs randomized individual participants, rather than clusters of participants. Overall, studies assigned 7,188 participants, ranging from 30 participants in one RCT (Brenner et al., 2000) to 570 participants in another (Kasper et al., 2008). Twenty-two studies did not report any information about a power calculation (Behnke et al., 2002; Bernhardt, Liske, and Ebeling, 1993; Bjerkenstedt, 2005 et al.; Fava et al., 2005; Gastpar, Singer, and Zeller, 2006; Hänsgen, Vesper, and Ploch, 1994; Harrer, Hübner, and Podzuweit, 1994; Kalb, Trautmann-Sponsel, and Kieser, 2001; Laakmann, Dienel, and Kieser, 1998; Lecrubier et al., 2002; Lenoir, Degenring, and Saller, 1999; Liu et al., 2010; Montgomery, Hübner, and Grigoleit, 2000; Moreno et al., 2005; Pakseresht et al., 2012; Philipp, Kohnen, and Hiller, 1999; Rahman et al., 2008; Schrader, Meier, and Brattström, 1998; Schrader, 2000; Uebelhack et al., 2004; Vorbach, Arnoldt, and Hubner, 1997; Witte et al., 1995), ten studies reported an *a priori* power calculation with targeted sample size achieved (Gastpar, Singer, and Zeller, 2005; Harrer et al., 1999; Kasper et al., 2006; Kasper et al., 2008; Mannel et al., 2010; Szegedi et al., 2005; van Gurp et al., 2002; Volz, Eberhardt, and Grill, 2000; Wheatley, 1997; Woelk, 2000), and three studies noted a post hoc analysis indicating insufficient power (Brenner et al., 2000; HDTSG, 2002; Shelton et al., 2001).

Setting

Four studies were conducted in the United States (Brenner et al., 2000; Fava et al., 2005; HDTSG, 2002; Shelton et al., 2001), 18 took place in Germany (Bernhardt, Liske, and Ebeling, 1993; Gastpar, Singer, and Zeller, 2005; Gastpar, Singer, and Zeller, 2006; Hänsgen, Vesper, and Ploch, 1994; Harrer, Hübner, and Podzuweit, 1994; Kalb, Trautmann-Sponsel, and Kieser, 2001; Kasper et al., 2006; Laakmann, Dienel, and Kieser, 1998; Mannel et al., 2010; Philipp, Kohnen, and Hiller, 1999; Schrader, Meier, and Brattström, 1998; Schrader, 2000; Szegedi et al., 2005; Uebelhack et al., 2004; Volz, Eberhardt, and Grill, 2000; Vorbach, Arnoldt, and Hubner, 1997; Woelk, 2000; Witte et al., 1995), one took place in both Germany and Sweden (Kasper et al., 2008), one took place in both Germany and Switzerland (Lenoir, Degenring, and Saller, 1999), two took place in the United Kingdom (Montgomery, Hübner, and Grigoleit, 2000; Wheatley, 1997), and one study each took place in France (Lecrubier et al., 2002), China (Liu et al., 2010), Brazil (Moreno et al., 2005), Iran (Pakseresht et al., 2012), Pakistan (Rahman et al., 2008), Canada (van Gurp et al., 2002), and Sweden (Bjerkenstedt et al., 2005). The country was not reported in two studies (Behnke et al., 2002; Harrer et al., 1999).

Twenty-six studies took place at multiple sites (Behnke et al., 2002; Bjerkenstedt et al., 2005; Fava et al., 2005; Gastpar, Singer, and Zeller, 2005; Gastpar, Singer, and Zeller, 2006; HDTSG, 2002; Hänsgen, Vesper, and Ploch, 1994; Harrer, Hübner, and Podzuweit, 1994; Harrer et al., 1999; Kalb, Trautmann-Sponsel, and Kieser, 2001; Kasper et al., 2006; Laakmann, Dienel, and Kieser, 1998; Lecrubier et al., 2002; Lenoir, Degenring, and Saller, 1999; Mannel et al., 2010; Montgomery, Hübner, and Grigoleit, 2000; Philipp, Kohnen, and Hiller, 1999; Schrader, Meier, and Brattström, 1998; Schrader, 2000; Shelton et al., 2001; Szegedi et al., 2005; Volz, Eberhardt, and Grill, 2000; Vorbach, Arnoldt, and Hubner, 1997; Wheatley, 1997; Woelk, 2000; Witte et al., 1995), while eight were at a single site (Bernhardt, Liske, and Ebeling, 1993; Brenner et al., 2000; Liu et al., 2010; Moreno et al., 2005; Pakseresht et al., 2012; Rahman et al., 2008; Uebelhack et al., 2004; van Gurp et al., 2002). The number of sites was not reported in one study (Kasper et al., 2008).

Participants

The age of participants ranged from 18–94 years. All studies included both male and female participants except for one, which did not provide information on gender (Montgomery, Hübner, and Grigoleit, 2000). The proportion of males ranged from 13 percent to 43 percent. Only one study included a comorbid mental health or medical disorder in more than three-quarters of its participants (unstable angina pectoris) (Liu et al., 2010).

Interventions

The total length of treatment with SJW ranged from four to 26 weeks. Ten RCTs specified the extract of SJW as LI 160 (0.3% hypericin, 1–4% hyperforin). Dosages given included 900

mg per day (Bjerkenstedt et al., 2005; Fava et al., 2005; Hänsgen, Vesper, and Ploch, 1994; Harrer, Hübner, and Podzuweit, 1994; Montgomery, Hübner, and Grigoleit, 2000; Wheatley, 1997), 600 mg per day (Mannel et al., 2010) 600–900 mg per day (Brenner et al., 2000), 900–1,500 mg per day (HDTSG, 2002), and 1,800 mg per day (Vorbach, Arnoldt, and Hubner, 1997). STW3-VI (0.2% hypericin, 2% hyperforin) with dosage of 900 mg per day was used in two studies (Gastpar, Singer, and Zeller, 2006; Uebelhack et al., 2004). WS 5570 (3–6% hyperforin, 0.1–0.3% hypericin) was used in four studies with dosages of 900 mg per day (Kasper et al., 2008; Lecrubier et al., 2002), 600–1,200 mg per day (Kasper et al., 2006), and 900–1,800 mg per day (Szegedi et al., 2005). Ze 117 (0.2% hyperforin) was used in three studies, with dosage of 500 mg per day (Schrader, Meier, and Brattström, 1998, Schrader, 2000; Woelk, 2000). We only identified one RCT for some of the extracts: STW3 (Gastpar, Singer, and Zeller, 2005), LoHyp-57 (Harrer et al., 1999), WS 5572 (5% hyperforin, 0.14% hypericin; Kalb, Trautmann-Sponsel, and Kieser, 2001), WS 5572 and WS 5573 (0.5% hyperforin, 0.14% hypericin; Laakmann, Dienel, and Kieser, 1998), D-0496 (hypericin; Volz, Eberhardt, and Grill, 2000), psychotonin forte (0.5% hypericin; Witte et al., 1995), and STEI 300 (2–3% hyperforin, 0.2–0.3% hypericin and pseudohypericin; Philipp, Kohnen, and Hiller, 1999). Two studies specified hypericin without further details (Bernhardt, Liske, and Ebeling, 1993; Lenoir, Degenring, and Saller, 1999). The rest of the RCTs stated the treatment as SJW, or Hypericum perforatum, but did not specify the extract used. We identified one RCT that utilized SJW as adjunctive therapy (Pakseresht et al., 2012), while the rest of the RCTs studied SJW as monotherapy.

Comparators

We found 12 two-arm RCTs providing data on treatment with SJW versus treatment with antidepressants (Behnke et al., 2002; Brenner et al., 2000; Fava et al., 2005; Gastpar, Singer, and Zeller, 2005; Harrer, Hübner, and Podzuweit, 1994; Harrer et al., 1999; Schrader, 2000; Szegedi et al., 2005; van Gurp et al., 2002; Vorbach, Arnoldt, and Hubner, 1997; Wheatley, 1997; Woelk, 2000). We found 12 two-arm RCTs comparing treatment with SJW to placebo (Hänsgen, Vesper, and Ploch, 1994; Kalb, Trautmann-Sponsel, and Kieser, 2001; Kasper et al., 2008; Lecrubier et al., 2002; Mannel et al., 2010; Montgomery, Hübner, and Grigoleit, 2000; Rahman et al., 2008; Schrader, 2000; Shelton et al., 2001; Uebelhack et al., 2004; Volz, Eberhardt, and Grill, 2000; Witte et al., 1995). Seven RCTs had three arms, comparing treatment with SJW to both antidepressant treatment and placebo (Bjerkenstedt et al., 2005; Fava et al., 2005; Gastpar, Singer, and Zeller, 2006; HDTSG, 2002; Moreno et al., 2005; Pakseresht et al., 2012; Philipp, Kohnen, and Hiller, 1999). Two RCTs had three arms, comparing two different doses or extracts of SJW with each other and placebo (Kasper et al., 2006; Laakmann, Dienel, and Kieser, 1998). One RCT compared three different doses of SJW (Lenoir, Degenring, and Saller, 1999) and one RCT compared SJW with antidepressant treatment and psychotherapy and a control agent (Liu et al., 2010). One RCT compared two different dosing schedules of SJW (Bernhardt, Liske, and Ebeling, 1993).

For subgroup analyses within antidepressant medication comparators, we differentiated selective serotonin reuptake inhibitor (SSRI), tricyclic antidepressants (imipramine, amitriptyline), and other (e.g., maprotiline, deanxit).

Outcomes

The majority of studies (Behnke et al., 2002; Bjerkenstedt et al., 2005; Brenner et al., 2000; Fava et al., 2005; Gastpar, Singer, and Zeller, 2005; Gastpar, Singer, and Zeller, 2006; Hänsgen, Vesper, and Ploch, 1994; Harrer, Hübner, and Podzuweit, 1994; Harrer et al., 1999; HDTSG, 2002; Kalb, Trautmann-Sponsel, and Kieser, 2001; Kasper et al., 2006; Laakmann, Dienel, and Kieser, 1998; Lecrubier et al., 2002; Liu et al., 2010; Mannel et al., 2010; Montgomery, 2000; Moreno et al., 2005; Philipp, Kohnen, and Hiller, 1999; Schrader, Meier, and Brattström, 1998; Schrader, 2000; Shelton et al., 2001; Szegedi et al., 2005; Uebelhack et al., 2004; Volz, Eberhardt, and Grill, 2000; Vorbach, Arnoldt, and Hubner, 1997; Wheatley, 1997; Witte et al., 1995; Woelk, 2000) reported on the number of responders to treatment. We found 25 RCTs measuring response to treatment in a continuous fashion (Behnke et al., 2002; Bjerkenstedt et al., 2005; Brenner et al., 2000; Fava et al., 2005; Gastpar, Singer, and Zeller, 2005; Gastpar, Singer, and Zeller, 2006; HDTSG, 2002; Kalb, Trautmann-Sponsel, and Kieser, 2001; Kasper et al., 2006; Kasper et al., 2008; Laakmann, Dienel, and Kieser, 1998; Lecrubier et al., 2002; Liu et al., 2010; Mannel et al., 2010; Pakseresht et al., 2012; Philipp, Kohnen, and Hiller, 1999; Schrader, Meier, and Brattström, 1998; Schrader, 2000; Shelton et al., 2001; Szegedi et al., 2005; Uebelhack et al., 2004; van Gurp et al., 2002; Volz, Eberhardt, and Grill, 2000; Vorbach, Arnoldt, and Hubner, 1997; Wheatley, 1997). We found 13 RCTs measuring remission from a major depressive episode (Behnke et al., 2002; Bjerkenstedt et al., 2005; Fava et al., 2005; Harrer, Hübner, and Podzuweit, 1994; Harrer et al., 1999; Kasper et al., 2006; Lecrubier et al., 2002; Moreno et al., 2005; Schrader, Meier, and Brattström, 1998; Shelton et al., 2001; Szegedi et al., 2005; Uebelhack et al., 2004; Witte et al., 1995). We found two RCTs measuring relapse to a major depressive episode (Kasper et al., 2008; Gastpar, Singer, and Zeller, 2005).

We found two RCTs that reported data on both physical and mental quality of life (Kasper et al., 2006; Philipp, Kohnen, and Hiller, 1999).

All but two included RCTs reported on the presence or the absence of adverse events, but studies varied greatly in their rigor of reporting data. For adverse events, we grouped clusters of symptoms as follows:

- Gastrointestinal/metabolic-nutritional: nausea, diarrhea, gastroenteritis, abdominal pain, and constipation.
- Neurological/nervous system: headache, dry mouth, dizziness, numbness, any sleep issue, fatigue, lethargy, asthenia, sweating, tremor, pain, restlessness, thirst, and forgetfulness.
- Skin/musculoskeletal: vascular disorders, palpitations, heart complaints, and syncope as cardiovascular. We grouped skin and appendage disorders, joint pain, muscle pain/aches, rash, skin and integumentary system, musculoskeletal and connective tissue system

disorders, skin and subcutaneous system disorders, muscle spasms, muscle or joint stiffness, allergic skin reactions, pruritis, exanthema, photosensitivity, and swelling.
- Psychiatric: psychiatric disorders and anxiety.
- Respiratory/infectious: cold symptoms, flu, upper respiratory tract infection (URTI), infections and infestations, sinusitis, bronchitis, common cold, respiratory, thoracic and mediastinal disorders, cough, and herpes labialis.
- Other organ systems: diseases of liver and bile duct, ear and labyrinth disorders, eye disorders, renal and urinary disorders, reproductive system and breast disorders, urinary problems, blurred vision, and frequent urination.
- Sexual dysfunction: sexual difficulties, sexual dysfunction, and anorgasmia.

Study Quality/Risk of Bias for Individual Included Studies

The risk of bias and study quality for each of the individual included studies can be found in Table 3.2. Ten studies obtained a "good" quality rating. Fourteen studies were judged to be of fair quality. These studies had completeness of reporting of outcome data but were unclear in some aspects of the methods (Bjerkenstedt et al., 2005; Brenner et al., 2000; Harrer et al., 1999; Kasper et al., 2006; Lecrubier et al., 2002; Philipp, Kohnen, and Hiller, 1999; Schrader, 2000; Szegedi et al., 2005; Kasper et al., 2008; van Gurp et al., 2002; Woelk, 2000; Moreno et al., 2005; HDTSG, 2002; Pakseresht et al., 2012). Eleven further studies were judged to be of poor quality. For eight of these studies, this was primarily due to issues with completeness of reporting outcome data, such as inadequate or missing ITT analysis and/or less than 80-percent follow-up (Gastpar, Singer, and Zeller, 2005; Vorbach, Arnoldt, and Hubner, 1997; Harrer, Hübner, and Podzuweit, 1994; Rahman et al., 2008; Lenoir, Degenring, and Saller, 1999; Behnke, 2002; Hänsgen, Vesper, and Ploch, 1994; Fava et al., 2005). Two studies were poor primarily due to lack of blinding (Liu et al., 2010; Bernhardt, Liske, and Ebeling, 1993), and one due to insufficient information in the publication (Montgomery, Hübner, and Grigoleit, 2000).

Random sequence generation. Fifteen studies had unclear selection bias because they did not report their random sequence generation method; 20 other studies reported adequate random sequence generation methods (e.g., computerized random number generator).

Allocation concealment. Twenty-three studies had unclear selection bias because they did not report their allocation concealment method, whereas 12 other studies did give a method of allocation concealment.

Blinding of participants and providers. Six studies had unclear selection bias because they did not report the method of ensuring blinding; 27 other studies reported adequate blinding methods, and two studies were considered high risk of blinding not ensured.

Blinding of outcome assessors. Twenty-eight studies had unclear risk of detection bias because they did not report whether outcome assessors were blind to participant intervention conditions. Five studies had low risk of bias, as the authors explicitly indicated that the outcome assessors were blind to intervention assignment, and two studies had high risk of bias, indicating assessors were not blinded.

Outcome data. Twenty-six studies had low risk of attrition bias; none had high risk, and nine were unclear.

Selective outcome reporting. Three studies had low risk of reporting bias because the authors provided a protocol for the study or an *a priori* trial registration entry. The rest of the studies had unclear risk of bias because one of these was not provided.

Other. All of the studies had an adequate comparator and either did not use treatment as usual or indicated that both study arms received treatment as usual for depression in addition to the study intervention. None of the studies were cross-over trials and therefore appropriate washout was not applicable.

Table 3.2. Study Quality/Risk of Bias for Individual Included Studies

							Other Biases			
Study	Random Sequence Generation (selection bias)	Allocation Concealment (selection bias)	Blinding of Participants and Personnel (performance bias)	Blinding of outcome Assessors (detection bias)	Completeness of Reporting Outcome Data (attrition bias)	Selective Outcome Reporting (reporting bias)	All Receive Treatment as Usual, Only Treatment Group Receives SJW (no placebo for controls)	Appropriate Washout Period or Exclusion of Individuals Taking Personal Supplements	Baseline Assessment, Appropriate Statistical Analysis, Conflict of Interest	USPSTF Quality Rating[a]
Behnke et al., 2002	Unclear risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Poor
Bernhardt, Liske, and Ebeling, 1993	Unclear risk	Unclear risk	High risk	High risk	Unclear risk	Unclear risk	Low risk	NA	Unclear risk	Poor
Bjerkenstedt et al., 2005	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	Low risk	NA	Low risk	Fair
Brenner et al., 2000	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Fair
Fava et al., 2005	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Poor
Gastpar, Singer, and Zeller, 2005	Low risk	Unclear risk	Low risk	Unclear risk	Unclear risk	Unclear risk	Low risk	NA	Low risk	Poor
Gastpar, Singer, and Zeller, 2006	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Good
HDTSG, 2002	Low risk	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low risk	NA	Low risk	Fair
Hänsgen, Vesper, and Ploch, 1994	Low risk	Low risk	Low risk	Unclear risk	Unclear risk	Unclear risk	Low risk	NA	Low risk	Poor
Harrer, Hübner, and Podzuweit, 1994	Low risk	Unclear risk	Low risk	Unclear risk	Unclear risk	Unclear risk	Low risk	NA	Low risk	Poor
Harrer et al., 1999	Unclear risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Fair

Study	Random Sequence Generation (selection bias)	Allocation Concealment (selection bias)	Blinding of Participants and Personnel (performance bias)	Blinding of outcome Assessors (detection bias)	Completeness of Reporting Outcome Data (attrition bias)	Selective Outcome Reporting (reporting bias)	Other Biases			USPSTF Quality Rating[a]
							All Receive Treatment as Usual, Only Treatment Group Receives SJW (no placebo for controls)	Appropriate Washout Period or Exclusion of Individuals Taking Personal Supplements	Baseline Assessment, Appropriate Statistical Analysis, Conflict of Interest	
Kalb, Trautmann-Sponsel, and Kieser, 2001	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Good
Kasper et al., 2006	Low risk	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low risk	NA	Low risk	Fair
Kasper et al., 2008	Low risk	Low risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	Low risk	Low risk	Fair
Laakmann, Dienel, and Kieser, 1998	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Good
Lecrubier et al., 2002	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Fair
Lenoir, Degenring, and Saller, 1999	Unclear risk	Low risk	Unclear risk	Unclear risk	Unclear risk	Unclear risk	Low risk	NA	Low risk	Poor
Liu et al., 2010	High risk	Unclear risk	High risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Poor
Mannel et al., 2010	Low risk	Low risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Good
Montgomery, Hübner, and Grigoleit, 2000	Unclear risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Unclear risk	Poor
Moreno et al., 2005	Unclear risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Fair
Pakseresht et al., 2012	Unclear risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Fair
Philipp, Kohnen, and Hiller, 1999	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Fair
Rahman et al., 2008	Low risk	Low risk	Low risk	Unclear risk	Unclear risk	Unclear risk	Low risk	NA	Low risk	Poor

Study	Random Sequence Generation (selection bias)	Allocation Concealment (selection bias)	Blinding of Participants and Personnel (performance bias)	Blinding of outcome Assessors (detection bias)	Completeness of Reporting Outcome Data (attrition bias)	Selective Outcome Reporting (reporting bias)	Other Biases			USPSTF Quality Rating[a]
							All Receive Treatment as Usual, Only Treatment Group Receives SJW (no placebo for controls)	Appropriate Washout Period or Exclusion of Individuals Taking Personal Supplements	Baseline Assessment, Appropriate Statistical Analysis, Conflict of Interest	
Schrader, Meier, and Brattström, 1998	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Good
Schrader, 2000	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Fair
Shelton et al., 2001	Low risk	Low risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Good
Szegedi et al., 2005	Low risk	Low risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Fair
Uebelhack et al., 2004	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Good
Volz, Eberhardt, and Grill, 2000	Unclear risk	Low risk	Low risk	Low risk	Low risk	Unclear risk	Low risk	NA	Low risk	Good
Vorbach, Arnoldt, and Hubner, 1997	Low risk	Unclear risk	Low risk	Unclear risk	Unclear risk	Unclear risk	Low risk	NA	Low risk	Poor
Wheatley, 1997	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Good
Witte et al., 1995	Unclear risk	Low risk	Low risk	High risk	Low risk	Unclear risk	Low risk	NA	Low risk	Good
Woelk, 2000	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	NA	Low risk	Fair
van Gurp et al., 2002	Low risk	Low risk	Low risk	Low risk	Low risk	Unclear risk	Low risk	NA	Low risk	Fair

NOTE: NA = not applicable; USPSTF = U.S. Preventive Services Task Force.

[a] The USPSTF criteria (U.S. Preventive Services Task Force, 2008) for study quality involve assessment of various factors related to the internal validity of the study. “Good” is the highest ranking, which involves comparable groups with low attrition, with outcomes being reliably and validly measured and analyzed. “Fair” is the next highest rating and involves studies with one or a few potential concerns (e.g., some though not major differences between groups exist at follow-up), though intention-to-treat analysis was performed. “Poor” is the lowest ranking and involves studies with one or more “fatal flaws” (e.g., no intention-to-treat analysis).

KQ 1: What Are the Efficacy and Safety of St. John's Wort, as an Adjunctive or Monotherapy, for Depressive Symptoms and Quality of Life in Adults with Major Depressive Disorder Compared with Placebo or Active Comparator?

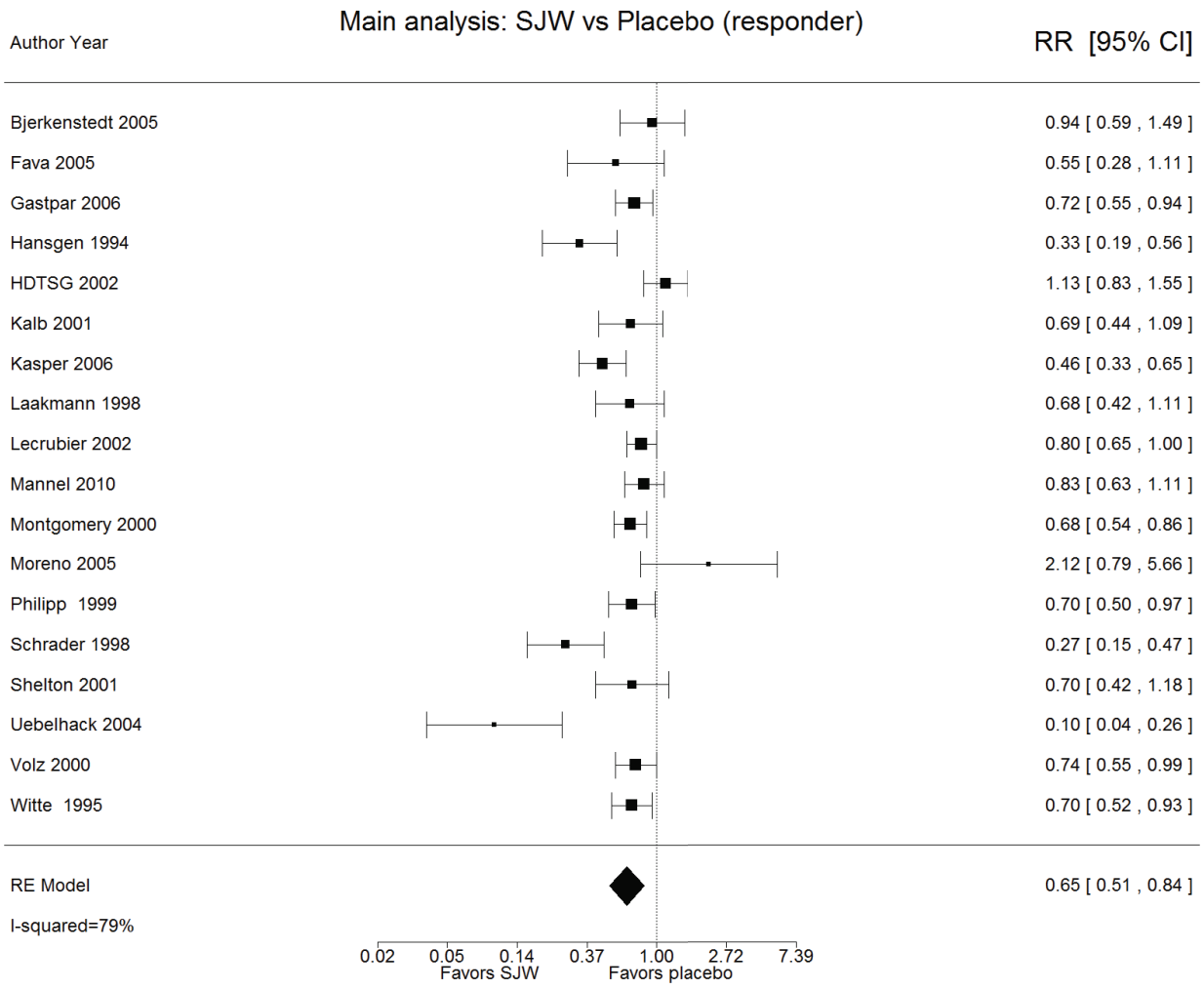
We identified 35 RCTs providing data on the efficacy of SJW and 34 RCTs addressing the presence or absence of adverse events. The effectiveness and safety compared with placebo comparators are documented below. The comparative effectiveness and safety of SJW are documented in KQ 1d and KQ 1e.

Included studies reported on a variety of depression outcome measures. Only two studies reported quality of life effect estimates. Studies varied in their approach to reporting safety.

Depression Treatment Responders

Eighteen RCTs reported the number of treatment responders per study arm comparing SJW with placebo. In the large majority of studies, treatment response was defined as a 50-percent decrease in HAMD scores. The median follow-up time was six weeks, with a range of four to 12 weeks. Although the large majority of individual studies indicated a positive trend in favor of SJW, many individual studies did not report statistically significant effects of SJW, and the size of the treatment effect varied somewhat across studies (see Figure 3.2). The pooled analysis across studies indicated a statistically significant treatment effect in favor of SJW (RR 0.65; CI 0.51, 0.84; 18 RCTs; I^2 79%). However, there was evidence of heterogeneity, and confidence intervals of some individual studies did not overlap, indicating nonexplained variance across study estimates.

Figure 3.2. St. John's Wort Versus Placebo, Treatment Responder



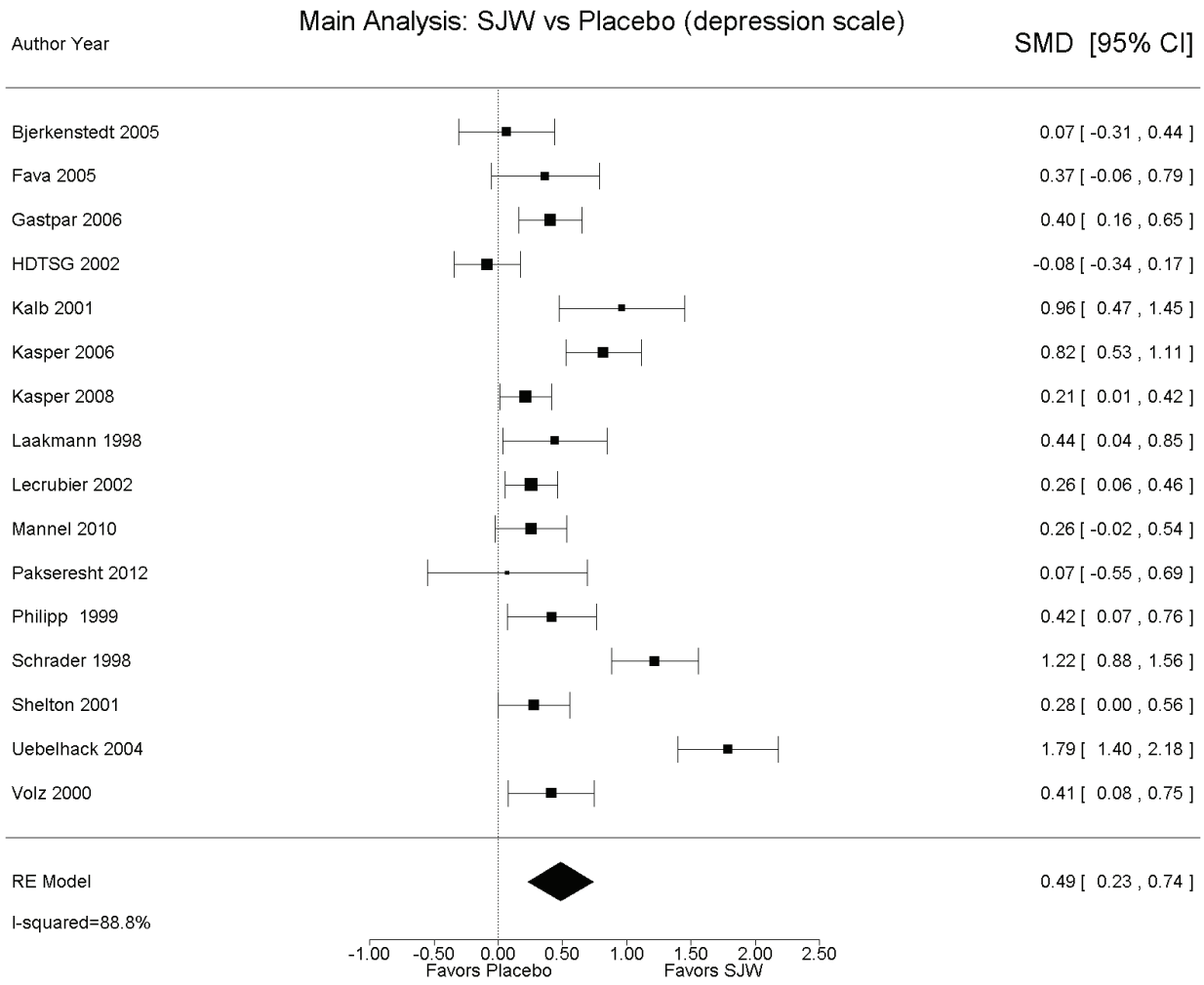
To determine whether the positive effect of SJW shown in the studies was primarily driven by poor methodological quality, we restricted the effectiveness analysis to studies determined to be at least fair or good. The sensitivity analysis showed very similar results for the number of participants showing a treatment response in favor of SJW over placebo when excluding poor quality studies (RR 0.68; CI 0.51, 0.91; 15 RCTs; I^2 80%). Heterogeneity was not reduced in this more selected study sample compared with the main analysis.

Depression Treatment Response Standardized Mean Differences

Sixteen RCTs provided data on continuous outcome scales assessing depression symptoms for both treatment arms; the large majority of studies used the HAMD to measure treatment

effects. Most individual studies reported treatment effects superior to placebo, but not all identified studies reported effects that were statistically different across treatment arms. The median follow-up time was six weeks and ranged from four to 32 weeks The pooled treatment estimate across studies indicated a statistically significant effect of SJW (SMD 0.49; CI 0.23, 0.74; 16 RCTs; I^2 89%); however, there was evidence of heterogeneity, as shown in Figure 3.3.

Figure 3.3. St. John's Wort Versus Placebo, Standardized Mean Differences

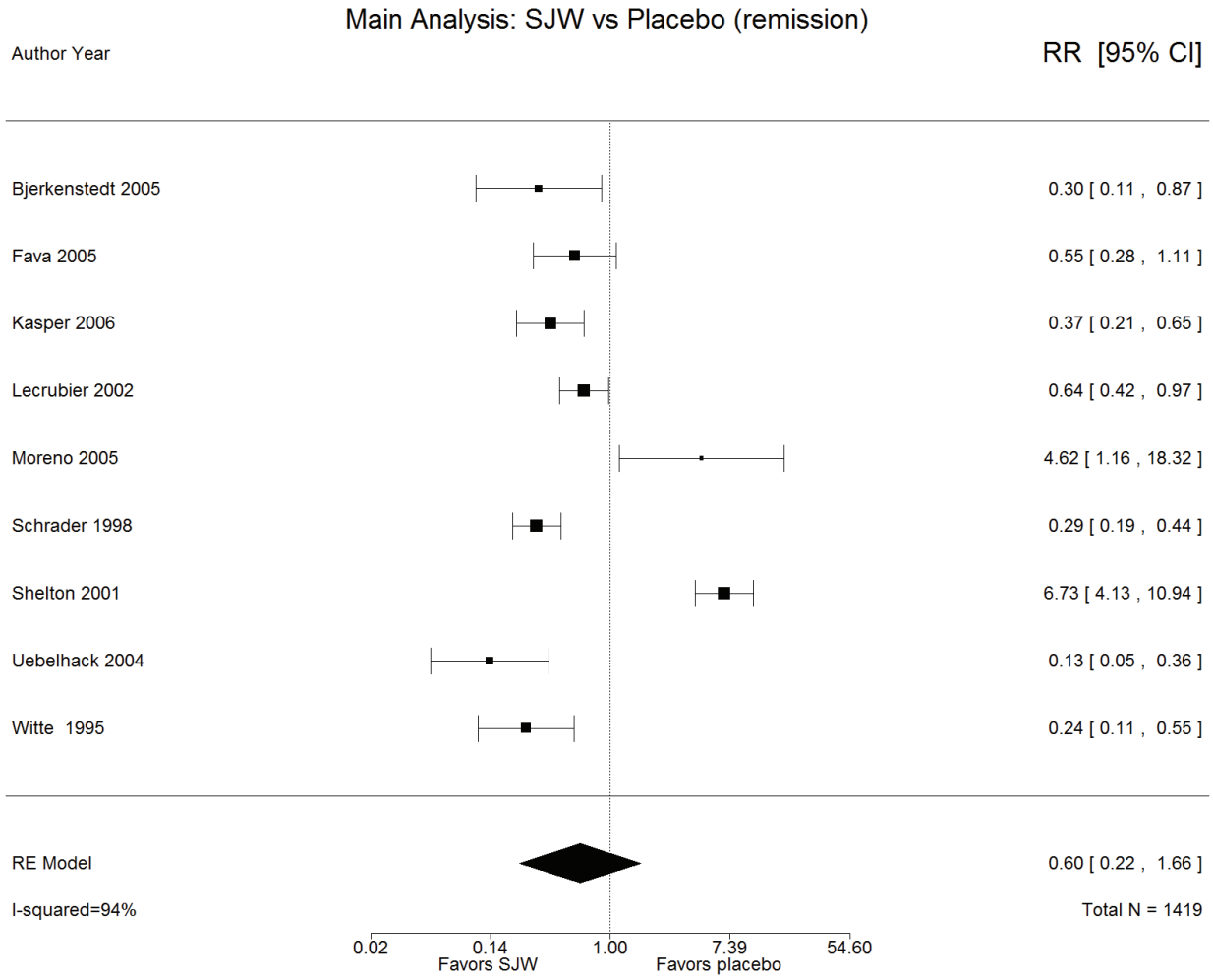


When excluding poor quality studies, pooled results were similar to the main analysis (SMD 0.50; CI 0.22, 0.77; 15 RCTs; I^2 90%), and heterogeneity continued to be very high.

Depression Remission

Nine RCTs reported on the number of participants in remission (i.e., not requiring treatment for depression anymore after the intervention). Although several individual studies reported a trend in favor of SJW, only half reported statistically significant effects. The pooled effect shows no statistically significant difference between the SJW and the placebo treatment arms across studies (RR 0.60; CI 0.22, 1.66; 9 RCTs; I^2 94%). The median follow-up time was six weeks (range four to 12 weeks) in the studies shown in Figure 3.4.

Figure 3.4. St. John's Wort Versus Placebo, Remission



The results were very similar when excluding all poor quality studies (RR 0.61; CI 0.19, 1.98; 8 RCTs; I^2 95%), and between-study heterogeneity was not reduced.

Depression Relapse

One RCT reported on relapse to a depressive episode for all participants randomized to SJW treatment or to placebo. The study did not indicate statistically significant difference between study arms (RR 1.42; CI 0.98, 2.06; 1 RCT).

Quality of Life

Two RCTs reported effects on health-related quality of life; both used the SF-36 to assess the outcome. Across studies there was a statistically significant effect for the mental component (SMD 0.48; CI 0.24, 0.73; 2 RCTs; I^2 0%) but not the physical component (SMD 0.28; CI −1.03, 0.47; 2 RCTs; I^2 0%).

Participants with Adverse Events

Thirteen studies reported the number of participants with adverse events per study arm. Across studies, there were no statistically significant differences between SJW and placebo treatment groups (OR 0.83; CI 0.62, 1.13; 13 RCTs). All studies included in this analysis were of fair or good quality. As a sensitivity analysis, we also compared the total number of reported adverse events per study arm. The analysis showed fewer adverse events in the SJW groups compared with placebo groups, but there was no statistically significant difference between groups (OR 0.75; est. CI 0.54, 75.48; 11 RCTs). This analysis has to be interpreted with caution, because individual participants may report more than one adverse event, and confidence intervals cannot be accurately computed.

Six studies reported serious adverse events with patients requiring hospitalization; pooled analyses showed a lower event rate in the SJW group but no statistically significant difference in events between study arms (OR 0.26; CI 0.04, 1.23; 6 RCTs). The result did not change with the exclusion of a poor quality study that reported no serious adverse events in either group. Of note, one study (van Gurp et al., 2002) with active control (sertraline) reported that a patient in the SJW group developed an acute manic reaction and was hospitalized. Hypertensive crisis was reported in one study (Szegedi et al., 2005), also an active control study (paroxetine). The study reported one event in 122 WS 5570 participants, but study authors determined it to be unrelated to the intervention because another cause was evident.

Individual Adverse Events

Fourteen RCTs reported on adverse events that were grouped as neurologic/nervous system events; pooled analyses showed statistically significantly more events in the SJW compared with placebo study arms (OR 1.56; CI 1.08, 3.32; 14 RCTs).

Fifteen studies reported the number of participants experiencing an adverse gastrointestinal/metabolic-nutritional event; pooled analyses showed no difference in the rate of events between SJW and placebo study arms (OR 1.08; CI 0.83, 1.41; 15 RCTs).

Ten studies reported on skin/musculoskeletal events; pooled analyses did not show statistically significantly more events in the SJW group (OR 0.98; CI 0.98, 2.21; 10 RCTs). Photosensitivity, specifically, was addressed in four RCTs. Kasper et al. (2006) reported two patients with increased sensitivity to sunlight and moderate sunburn in 250 patients taking SW 5570 (dosage arm not specified). Pakseresht et al. (2012) reported photosensitivity for three of 20 SJW patients taking SW 5570. Kasper et al. (2008) also addressed photosensitivity and found three cases in the placebo group compared with one in the WS 5570 group. Rahman et al. (2008) reported four cases of photosensitivity in both study arms (56 patients each in the SJW and placebo group). There was no statistically significant difference across studies reporting on photosensitivity (OR 1.10; CI 0.36, 3.56; 4 RCTs). Of note, three included studies excluded participants with known photosensitivity from the trial (Gastpar, Singer, and Zeller , 2005; Gastpar, Singer, and Zeller, 2006; Uebelhack et al., 2004).

Five studies reported on other organ system (e.g., eye, ear, liver, renal, reproductive) events; pooled analyses showed statistically significantly more events in the SJW compared with placebo study arms (OR 1.87; CI 1.08, 3.32; 5 RCTs).

Seven studies reported on respiratory/infectious events; pooled analyses showed a trend for more events in the SJW group but no statistically significant difference in event rates between study arms (OR 1.48; CI 0.95, 2.33; 7 RCTs).

Four studies reported the number of cardiovascular adverse events; pooled analyses showed a higher frequency of events in the SJW group (ten events in 493 patients versus one in 266 patients), but the difference was not statistically significant between arms (OR 6.81; CI 0.92, 304.08; 4 RCTs).

Three studies reported on psychiatric adverse events; pooled analyses showed a trend for more events in the SJW group but, given the rarity of events, no statistically significant difference in event rates between study arms (OR 1.61; CI 0.34, 10.21; 3 RCTs).

Two RCTs reported on sexual dysfunction events; we found statistically significantly more events in the SJW group, but the pooled analyses showed no statistically significantly different effect between arms (OR 1.92; CI 0.94, 4.00; 2 RCTs).

Regarding other specific adverse events that have been associated with SJW in the literature, thyroid stimulating hormone associated events were not addressed in any of the included RCTs.

Drug interactions with concomitant medication were addressed in three included studies (Gastpar, Singer, and Zeller, 2005; Gastpar, Singer, and Zeller, 2006; Manell et al., 2011), and all reported that no interactions were observed; however, all but one included RCT investigated SJW as monotherapy with no other systematic concomitant depression treatment.

Although pregnancy was not mentioned in any of the studies, 15 included studies did not specify pregnancy in their exclusion criteria. None of the included studies reported peri- or postnatal adverse events.

Study Characteristic Moderators and Risk of Bias

Meta-regressions investigating the effect of study quality on effect sizes using the overall quality rating indicated no association for the outcome responder (p=0.321), depression scale scores (p=0.195), or remission (p=0.956).

Although we searched the international literature without language restriction, a large proportion of included studies were conducted in Germany. To explore whether effect estimates are associated with the study setting, we differentiated German and non-German studies. A meta-regression found no indication that effect sizes in the outcome number of responders (p=0.078) or number of patients with adverse events (p=0.95) are associated with the setting. However, results for the continuous outcome (change in depression rating scales) indicated a systematic effect (p=0.012), as did the results for the outcome depression remission (p=0.058), with results in German studies reporting a stronger effect of SJW than non-German studies. In a subgroup analysis excluding all German studies, the effect for treatment responders in favor of SJW was smaller and was not statistically significant (RR 0.70; CI 0.42, 1.10; 7 RCTs; I^2 44%). The depression scale score analysis still showed a statistically significant effect in favor of SJW (SMD 0.18; CI 0.04, 0.31; 7 RCTs; I^2 15%). Effects for the outcome remission in the non-German samples showed no statistically significant difference between SJW and placebo (RR 1.25; CI 0.23, 6.93; 5 RCTs; I^2 94%), but there was considerable heterogeneity between studies.

Tests for publication bias for the outcome number of responders (Egger test p=0.142, Begg test p=0.069), depression scale scores (Egger test p=0.434, Begg test p=0.064), depression remission (Egger test p=0.920, Begg test p=0.893), number of participants with adverse events (Egger test p=0.555, Begg test p=0.324), or across specific adverse event categories (Egger test p=0.509, Begg test p=0.350) were not statistically significant. Estimating the treatment effect for the number of responders and depression scale scores using the trim-and-fill method did not result in different treatment estimates, and no hypothetical studies were added (responder outcome standard error [SE] 2.18, depression scale score SE 2.00). Of note, the review was not limited to peer-reviewed published articles and included grey literature, such as conference abstracts.

KQ 1a: Is SJW More Effective as Monotherapy Than as an Adjunctive Therapy?

There were 35 RCTs providing data on SJW as monotherapy and only one RCT providing data on SJW as specifically adjunctive therapy. Hence, the presented evidence in this review is primarily based on monotherapy studies.

Monotherapy

With regard to depression measures, the responder, remission, and relapse data presented previously were entirely based on monotherapy studies. For the continuous depression outcome,

analysis results did not change after excluding the adjunctive therapy study (Pakseresht et al., 2012) with pooled results estimating SMDs of 0.51 (CI 0.24, 0.78; 15 RCTs, I^2 90%) in favor of SJW compared with placebo. Excluding the adjunctive therapy study did also not reduce the considerable heterogeneity between studies.

The only adjunctive therapy study did not report on the outcome quality of life; hence, treatment estimates of SJW given as monotherapy are as reported above.

The adverse event analysis presented in the previous section applies to this subsection, because the data are based on monotherapy studies.

Some included studies allowed patients to continue treatments they were already using, others did not report on potential co-interventions, while others explicitly excluded the use of psychoactive medications for the duration of the SJW treatment trial.

When restricting the responder analysis to studies that explicitly excluded patients on antidepressants, treatment effect estimates were unchanged (RR 0.65; CI 0.49, 0.86; 16 RCTs; I^2 81%). Heterogeneity was somewhat reduced but still considerably high. The corresponding analysis for continuous outcomes was also similar to the main analysis (SMD 0.52; CI 0.22, 0.82; 13 RCTs; I^2 90%), and heterogeneity was not reduced. Estimates for the outcome remission indicated a somewhat smaller effect in studies that explicitly did not allow patients to continue antidepressant use, but effects were still statistically significantly in favor of SJW (RR 0.71; CI 0.18, 2.78; 7 RCTs; I^2 96%), with no evidence of reduced heterogeneity. The relapse data are unchanged given that the only contributing RCT did not allow antidepressant use (RR 1.42; CI 0.98, 2.06; 1 RCT).

Quality of life effect estimates continued to indicate a significant effect for the mental component (SMD 0.35; CI 0.01, 0.70; 1 RCT) and the study that excluded all patients on antidepressants also reported a positive effect on the physical component of quality of life in favor of SJW compared with placebo (SMD 0.46; CI 0.11, 0.81; 1 RCT).

In this subgroup of studies, the number of participants with adverse events and the number of events were identical with the main analysis, because only monotherapy studies contributed to it. There was a statistically significant effect for serious adverse events in favor of SJW because all reported events occurred in placebo arms (OR 0.00; CI 0.00, 0.59; 5 RCTs). Analyses for individual adverse events showed somewhat more cardiovascular events (OR 2.17; CI 0.20, 111.24; 2 RCTs) but no difference in gastrointestinal/metabolic-nutritional (OR 0.97; CI 0.70, 1.34; 13 RCTs), neurologic/nervous system (OR 1.33; CI 0.98, 1.82; 11 RCTs), respiratory/infectious (OR 1.44; CI 0.89, 2.37; 6 RCTs), skin/musculoskeletal (OR 1.46; CI 0.91, 2.37; 6 RCTs), other organ (e.g., eye, ear, liver, renal, reproductive) systems (OR 1.78; CI 0.99, 3.24: 4 RCTs), or sexual dysfunction events (OR 1.92; CI 0.94, 4.00; 2 RCTs).

Adjunctive Therapy

Only one included study gave SJW systematically adjunctive to another intervention. In the study (Pakseresht et al., 2012), patients were randomized to receive a tricyclic antidepressant

(nortiptyline 75–100 mg, imipramine and amitriptyline 100–150 mg daily) and either perforan pills (providing 300 mcg hypericin) or placebo, three times daily. Both groups showed improvement over six weeks; the mean Beck score did not suggest a different treatment effect between patients receiving SJW and antidepressants and/or placebo and antidepressants (SMD 0.07; CI −0.55, 0.69).

The study did not report on quality of life.

The study reported that gastrointestinal complications in the SJW group were significantly lower than those of the placebo group, but three patients developed mild photosensitivity.

KQ 1b: Is There a Difference in Efficacy, Depending on the Amount and Type of Extract of SJW Used?

We identified four head-to-head trials, one RCT comparing two different extracts of SJW and three RCTs comparing different dosages. Included studies most frequently used the extract LI 160 (11 RCTs). Subgroup analyses were performed for LI 160, WS 5570, Ze 117, and STW3-VI (i.e., all extracts were used in more than one study). Information regarding the content of the preparations was extracted from the original study authors, based on the manufacturers' specifications.

Head-to-Head Comparisons

In the extract comparison study (Laakmann, Dienel, and Kieser, 1998), 147 patients received either placebo, Hypericum extract WS 5573 (300 mg with a content of 0.5% hyperforin) or Hypericum extract WS 5572 (300 mg with a content of 5% hyperforin); the authors stated that the manufacturing process was identical for both preparations and differed according to the fingerprint chromatogram only with regard to the hyperforin content. After the 42-day treatment period, there was no statistically significantly difference based on the outcome responder or SMDs in depression scale scores, but the authors reported that the monotonic trend indicated superiority of WS 5572 over WS 5573 (p=0.017). The study did not report on quality of life. The authors noted that the incidence of adverse events was lowest in the WS 5572 group (0.35 events per patient) compared with placebo (0.47 events per patient) and WS 5573 (0.49 events per patient). Adverse events were reported in 24 of 49 patients on WS 5573, 23 of 49 patients on placebo, and 17 of 49 patients on WS 5572. Adverse events included headache, bronchitis, flu-like symptoms, cough, infection, and herpes labialis.

An RCT of 332 patients (Kasper et al., 2006) comparing WS 5570 600 mg per day versus 1,200 mg per day reported that the HAMD total scores decreased over six weeks by 11.6 (standard deviation [SD] 6.4) points in the patients taking WS 5570 600 mg per day, by 10.8 (SD 7.3) points in the patients taking WS 5570 1200mg per day, and 6.0 (SD 8.1) points in those taking placebo. The differences between extracts were not statistically significantly different for the depression outcome treatment responders, SMDs, or remission. There was also no difference

in the mental or physical component of quality of life. Adverse events were reported in 49 of 123 patients on 600 mg per day, 50 of 127 patients on 1,200 mg per day, and 22 out of 82 patients on placebo. Adverse events included ear and labyrinth disorders, eye disorders, gastrointestinal disorders, general disorders, administration site conditions, infections and infestations, injury, poisoning and procedural complications, investigations, metabolism and nutrition disorders, musculoskeletal and connective tissue disorders, nervous system disorders, psychiatric disorders, renal and urinary disorders, reproductive and breast disorders, respiratory thoracic and mediastinal disorders, skin and subcutaneous disorders, vascular disorders, serious adverse events, and adverse events potentially related to the treatment.

A second RCT comparing dosages of hypericin had 348 patients who took either 0.17 mg, 0.33 mg, or 1 mg of hypericin per day. Although the related efficacy was about 4-percent better in the highest dose group than the lowest dose, the three-factor analysis of variance showed no significant differences, with all patients showing a 50-percent reduction in the HAMD-17 (one version of HAMD) score at the end of six weeks of treatment. The HAMD-17 scores started at 16.4–16.9 points and decreased to 8.0–8.7 points overall with response rates of 62 percent of those taking 0.17 mg, 65 percent of those taking 0.33 mg, and 68 percent of those taking 1 mg (Lenoir, Degenring, and Saller, 1999). The study did not report on quality of life. Adverse events occurred in 40 out of 83 patients on 1 mg, 25 out of 90 patients on 0.33 mg, and 17 out of 87 patients on 0.17 mg (not statistically significantly different between groups). Adverse events involved skin, skeleton/muscles, nerves, psyche, gastrointestinal tract, liver/biliary system, cardiovascular system, airways/lungs, blood, kidneys/urinary tract, reproductive organs, neoplasms, and organism as a whole.

One RCT (Bernhardt, Liske, and Ebeling, 1993) compared different dosing schedules giving 55 participants with mild to moderate depression three Esbericum capsules (0.75 mg hypericin) either in the morning and at lunch time or three times a day. There was a trend favoring the trice-a-day schedule, but HAMD scores did not statistically significantly differ between schedules. The study did not report on quality of life. Adverse events were not addressed (the study was only published as a conference abstract).

LI 160 Subgroup

Five RCTs testing the extract LI 160 (standardized content 0.3% hypericin and 1–4% hyperforin) compared with placebo reported on the number of treatment responders. The treatment effect estimate was identical to the main responder analysis; however, in this subgroup, the pooled effect was not statistically significant (RR 0.66; CI 0.40, 1.06; 5 RCTs; I^2 72%). Heterogeneity was reduced when restricting to the extract LI 160 but remained considerable.

Three of the RCTs testing LI 160 also reported sufficient data to compute the SMD compared with placebo. The treatment effect estimate was smaller and not statistically significant compared with placebo (SMD 0.23; CI −0.10, 0.56; 3 RCTs; I^2 0%). There was no evidence of statistical heterogeneity.

Two of these RCTs measured remission; one reported a statistically significant treatment effect, the other one did not. There was no evidence of heterogeneity, but the width of the confidence interval did not suggest a meaningful pooled effect (CI −9.94, 15.79; 2 RCTs, I^2 0%).

None of the studies in this subgroup reported on quality of life.

There was no difference in the number of patients with adverse events and the total number of adverse events in the LI 160 subgroup that reported these outcomes (OR 1.37; CI 0.06, 33.62; 2 RCTs; and OR 1.61; CI 0.26, 9.92; 2 RCTs). Two studies reported the number of serious adverse events; the one reported event was associated with the control group.

Specific adverse events reported in more than one RCT in this subgroup included neurologic/nervous system events, and the pooled analysis showed statistically significantly more events in the LI 160 group compared with placebo (OR 2.22; CI 1.44, 3.44; 5 RCTs). The number of events classified as other organ systems (e.g., eye, ear, liver, renal, reproductive system) was also statistically significantly higher in the LI 160 group (OR 2.72; CI 1.31, 5.88; 2 RCTs). There was no difference between gastrointestinal/metabolic-nutritional events across study arms (OR 0.97; CI 0.64, 1.49; 5 RCTs). Two studies suggested somewhat more respiratory/infectious events in the LI 160 group, but there was no statistically significant difference between study arms (OR 1.88; CI 0.77, 4.78; 2 RCTs). Four studies suggested a somewhat higher rate in skin/musculoskeletal adverse events in the LI 160 group, but there were no statistically significant differences between study arms (OR 1.64; CI 0.96, 2.84; 4 RCTs). Across two studies, more sexual dysfunction adverse events were reported in LI 160 groups (28 of 212 versus 17 of 216), but there was no statistically significant difference between study arms (OR 1.92; CI 0.94, 4.00; 2 RCTs).

STW3-VI Subgroup

Two studies tested the effect of the extract STW3-VI (0.2% hypericin, 2% hyperforin) compared with placebo. Both studies reported statistically significant results for the number of responders (RR 0.72; CI 0.55, 0.94; and RR 0.10; CI 0.04, 0.26), but treatment effect estimates varied greatly, suggesting that a pooled analysis is not meaningful.

The studies also reported statistically significant SMDs compared with placebo (SMD 0.40; CI 0.16, 0.65; and SMD 1.79; CI 1.40, 2.18), but again, the pooled analyses showed extremely wide confidence intervals.

One of the studies (Uebelhack et al., 2004) also reported on the outcome remission indicating superiority of STW3-VI compared with placebo (RR 0.13; CI 0.05, 0.36; 1 RCT).

None of the STW3-VI studies reported on quality of life.

One RCT reported a serious adverse event; the event occurred in the control group.

Two RCTs in this subgroup reported on gastrointestinal/metabolic-nutritional event rates; there were fewer events in the STW3-VI group compared with placebo, but there was no statistically significant difference between arms (OR 0.60; 0.26, 1.33; 2 RCTs).

WS 5570 Subgroup

Two studies tested the effect of the extract WS 5570 (3–6% hyperforin, 0.1–0.3% hypericin) compared with placebo. One reported statistically significant results for the number of responders (RR 0.46; CI 0.33, 0.65), the other one approaching significance (RR 0.80; CI 0.65, 1), but treatment effect estimates varied greatly, suggesting that a pooled analysis is not meaningful.

Three RCTs reported sufficient data for SMDs. All individual studies were positive (SMD 0.82; CI 0.53, 1.11; SMD 0.21; CI 0.01, 0.42; and SMD 0.26; CI 0.06, 0.46); however, the pooled analysis in this subgroup analysis was not statistically significantly different from placebo (SMD 0.42; CI −0.41, 1.24).

Two studies reported on the outcome remission; both studies reported statistically significant effects of WS 5570 compared with placebo (RR 0.37; CI 0.21, 0.65; and RR 0.64; CI 0.42, 0.97), but the difference in treatment estimates and wide confidence intervals in the pooled analysis did not suggest that pooling is meaningful.

One of the WS 5570 RCTs (Kasper et al., 2008) reported on relapse but did not find a statistically significant effect compared with placebo (RR 0.70; CI 0.48, 1.02; 1 RCT).

The same study also reported on quality of life, with positive effects for the mental component (SMD 0.50; CI 0.22, 0.78) but not the physical component (SMD 0.23; CI −0.05, 0.51) compared with placebo.

One RCT (Kasper et al., 2006) reported three serious adverse events in the WS 5570 groups, with no event in the placebo group (tendon rupture attributable to accidental injury in the 600 mg per day arm; depression aggravation and acute stress disorder, attributable to the underlying disease and not tolerability issues in the 1200 mg per day arm). Two RCTs in this subgroup reported on gastrointestinal/metabolic-nutritional event rates; there were more events in the WS 5570 group compared with placebo, but there was no statistically significant difference between arms (OR 1.30; CI 0.76, 2.28; 2 RCTs). Studies reporting on neurologic/nervous system events showed no difference between WS 5570 and placebo (OR 1.01; CI 0.48, 2.18; 2 RCTs). There was also no statistically significant difference in skin/musculoskeletal adverse events across arms (OR 0.67; CI 0.26, 1.89; 2 RCTs).

Ze 117 Subgroup

Only one RCT (Schrader, Meier, and Brattström, 1998) using Ze 117 (0.2% hyperforin) reported on the outcome number of treatment responders favoring SJW over placebo (RR 0.27; CI 0.15, 0.47; 1 RCT). The same study also reported a positive effect using a continuous outcome (SMD 1.22; CI 0.88, 1.56; 1 RCT). Results for the number of patients in remission were also positive (RR 0.29; CI 0.19, 0.44).

The study did not report on quality of life.

Six out of 81 participants in the Ze 117 group and five out of 81 in the placebo group experienced an adverse event. In the Ze 117 group, this included two patients with moderate abdominal pain, one with moderate diarrhea, one with moderate melancholia, one with moderate acute deterioration, and one with mild dry mouth. In the placebo group, this included three with moderate abdominal pain, one with mild paresthesia, and one with a serious adverse event (severe syncope).

WS 5572 Subgroup

Two RCTs (Kalb, Trautmann-Sponsel, and Kieser, 2001; Laakmann, Dienel, and Kieser, 1998) reported on the number of responders after treatment with WS 5572 (5–6% hyperforin, 0.12–0.28% hypericin). The pooled result showed a statistically significant effect favoring WS 5572 over placebo (RR 0.69; CI 0.63, 0.74; 2 RCTs; I^2 0%), with no evidence of heterogeneity. The treatment effect using a continuous outcome was statistically significantly different from placebo in both studies (SMD 0.96; CI 0.47, 1.45; SMD 0.44; CI 0.04, 0.85), but treatment effect sizes varied and the pooled effect estimate showed confidence intervals too wide to suggest a meaningful pooled summary.

The RCTs testing WS 5572 did not report on quality of life.

In one of the RCTs (Kalb, Trautmann-Sponsel, and Kieser, 2001), there were three adverse events in the 34 patients in the WS 5572 group and two in the 35 placebo group participants. In the WS 5572 group, these included sinusitis, bronchitis, and the common cold. In the placebo group, these included bronchitis and gastroenteritis. In the other RCT (Laakmann, Dienel, and Kieser, 1998), 14 of 49 participants in the WS 5573 group, 14 of 49 in the WS 5572 group, and 15 of 49 in the placebo group experienced an adverse event. In the WS 5573 group, these included three patients with bronchitis, two with flu-like symptoms, two with cough, and one with infection. In the WS 5572 group, these included one patient with bronchitis. In the placebo group, these included five patients with headache, three with bronchitis, one with flu-like symptoms, one with cough, two with infection, and two with herpes labialis. Across studies, there was a lower rate of respiratory/infectious adverse events, but differences between arms were not significant (OR 0.56; CI 0.12, 2.30; 2 RCTs).

Meta-Regression for Extracts

We also performed a meta-regression for the dataset to identify whether treatment effects were associated with the type of extract used in the study. The analysis differentiated the subgroups LI 160, STW3-VI, WS 5570, Ze 117, WS 5572, and other. The meta-regression did not suggest that treatment estimates differ by extract when investigating the outcome treatment response (p=0.347), SMD (p=0.127), or remission (p=0.371).

KQ 1c: Is There a Difference in Efficacy, Depending on the Type of MDD (i.e., Mild, Moderate, Severe)?

The majority of studies included patients with mild depression, as well as patients with moderate depression.

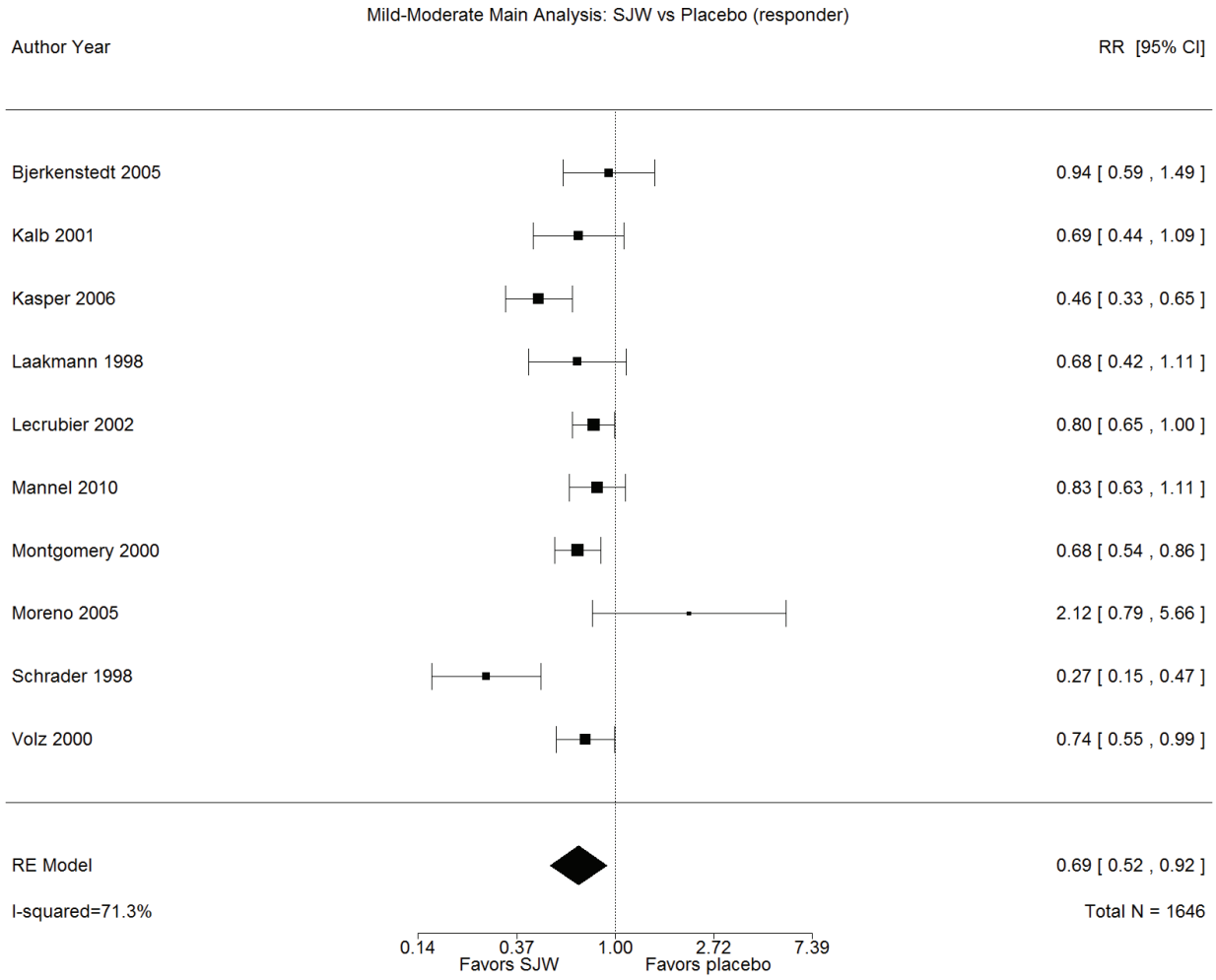
Mild Depression

We did not identify any study that was exclusively in patients with mild depression.

Mild-Moderate Depression

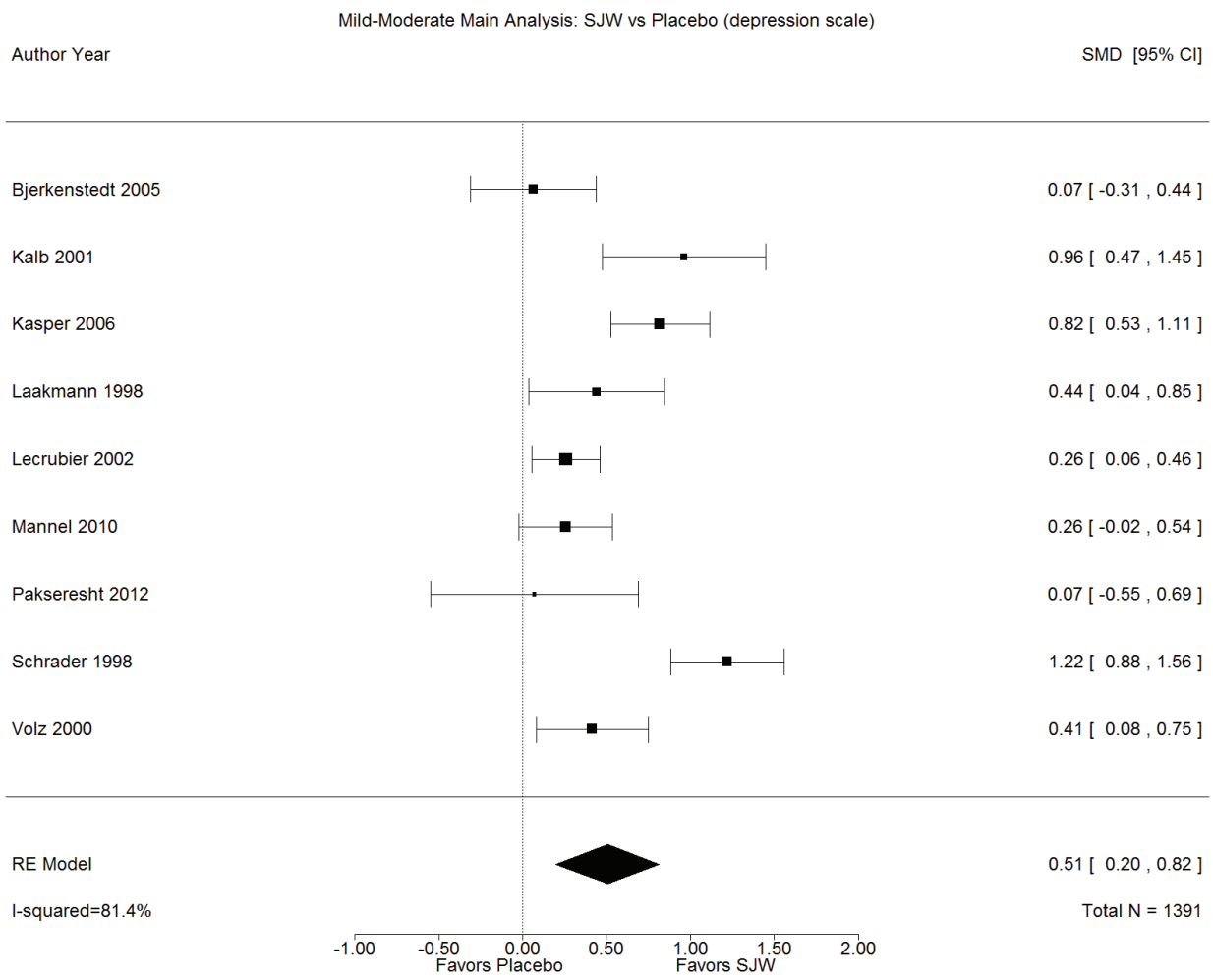
Twelve included studies reported on samples that included participants with either mild or moderate depression, making this the largest depression severity subgroup within the review dataset. Ten RCTs reported on the number of responders (RR 0.69; CI 0.52, 0.92; 10 RCTs; I^2 71%). The point estimate favoring SJW over placebo was similar to the main analysis that included all studies. The analysis of this selected patient subset still showed evidence of heterogeneity, as shown in Figure 3.5.

Figure 3.5. St. John's Wort Versus Placebo, Treatment Responder for Mild-Moderate Depression



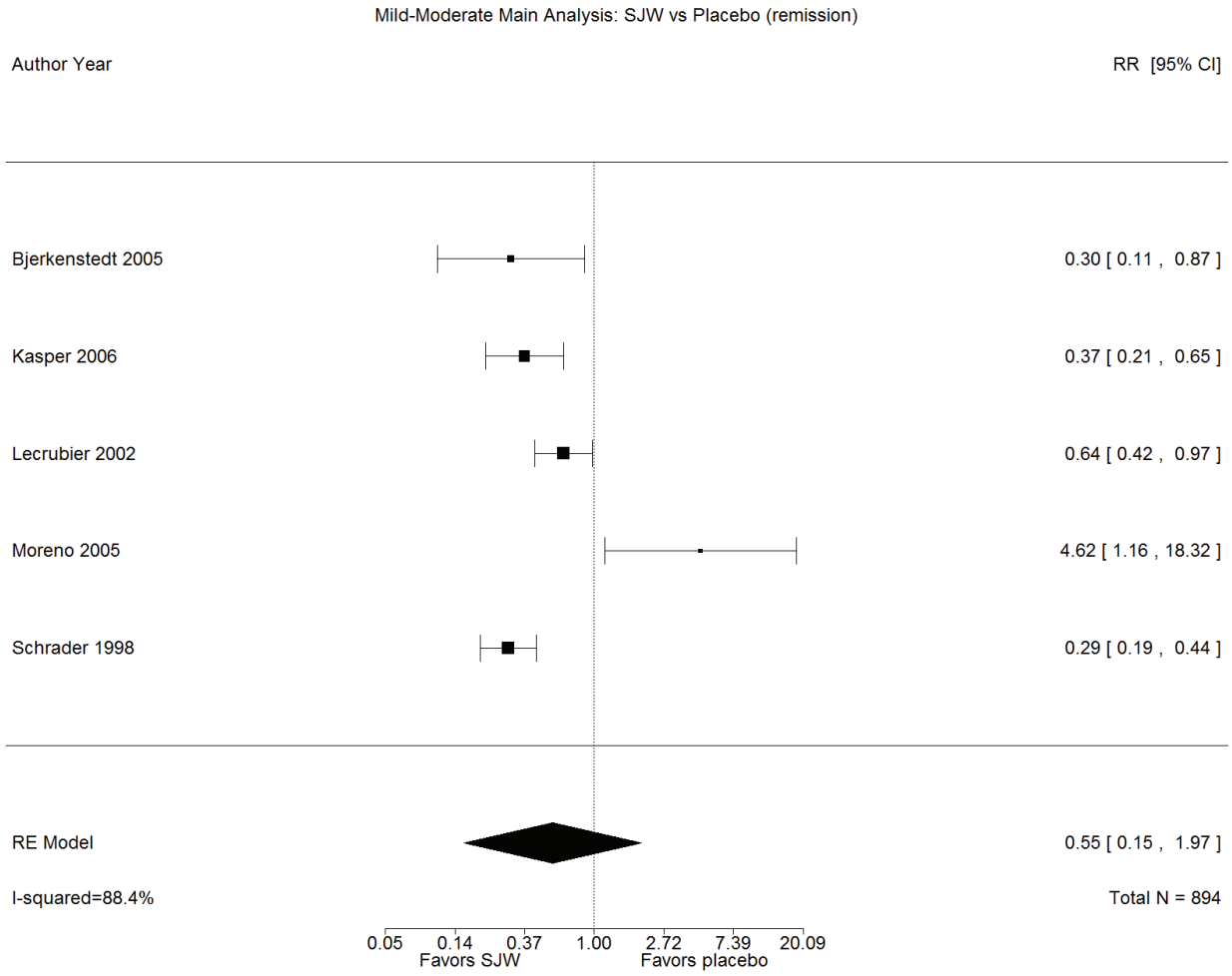
There was a statistically significant effect of SJW compared with placebo that was very similar to the main pooled analysis for depression measured as a continuous outcome (SMD 0.51; CI 0.20, 0.82; 9 RCTs; I^2=81%), as shown in the Figure 3.6.

Figure 3.6. St. John's Wort Versus Placebo, Standardized Mean Differences for Mild-Moderate Depression



Five of the RCTs studying mild and moderate depression patient samples reported on the outcome remission. In this subgroup, there were also no statistically significant differences between treatment and placebo groups (RR 0.55; CI 0.15, 1.97; 5 RCTs; I^2 88%), and heterogeneity was reduced compared with the main analysis, but was still considerable (see Figure 3.7).

Figure 3.7. John's Wort Versus Placebo on Remission for Mild-Moderate Depression



One RCT (Kasper et al., 2006) provided data on quality of life for a participant pool of mild and moderate depression. The study reported a statistically significant effect for the mental component (SMD 0.50; CI 0.22, 0.78) but not the physical component (SMD 0.23; CI 0.51, 0.05) of the SF-36 compared with placebo.

Seven RCTs reported the number of participants with adverse events in each study arm. Individual studies sometimes favored SJW and sometimes placebo arms; across studies, there was no statistically significant difference between arms (OR 1.10; 0.71, 1.71; 7 RCTs). The sensitivity analysis for the total number of adverse events in the study arms also indicated no significant differences between arms (OR 1.31; est. CI 0.86, 2.00; 7 RCTs). Four RCTs reported on a serious adverse event; across studies, there was a somewhat lower rate of events in the SJW

arm in this subgroup, but there was no statistically significant difference between groups (OR 0.79; CI 0.08, 10.24; 4 RCTs).

In this subgroup, there were somewhat more cardiovascular events in SJW groups compared with placebo groups, but due to the rarity of the event and the small number of studies reporting on the specific adverse event, there was no statistically significant difference between study arms (OR 4.86; CI 0.77, 1.70; 2 RCTs). There were also no statistically significant differences for gastrointestinal/metabolic-nutritional (OR 1.14; CI 0.77, 1.70; 8 RCTs), neurologic/nervous system (OR 1.45; CI 0.94, 2.25; 7 RCTs), respiratory/infectious (OR 1.75; CI 0.86, 3.76; 5 RCTs); skin/musculoskeletal (OR 1.73; CI 0.88, 3.51; 6 RCTs), other organ systems (e.g., eye, ear, liver, renal, reproductive) (OR 1.76; CI 0.59, 5.84; 3 RCTs), or psychiatric (OR 1.61; CI 0.34, 10.21; 3 RCTs) adverse events.

Moderate Depression

Of the RCTs providing results for participants with only moderate depression, three compared SJW with placebo (Gastpar, Singer, and Zeller, 2006; Philipp, Kohnen, and Hiller, 1999; Uebelhack et al., 2004). All three studies showed a statistically significant effect compared with placebo in terms of the number of responders (RR 0.72; CI 0.55, 0.94; RR 0.70; CI 0.50, 0.97; RR 0.10; CI 0.04, 0.26), but treatment effect size estimates varied widely and the pooled analysis was not statistically significant with a confidence interval crossing one (RR 0.40; CI 0.03, 6.23; 3 RCTs; I^2 96%), and there was evidence of considerable heterogeneity.

All three studies also showed a statistically significant effect compared with placebo in a continuous depression outcome (SMD 0.40; CI 0.16, 0.65; SMD 0.42; CI 0.07, 0.76; SMD 1.79: CI 1.40, 2.18), but treatment effect size estimates varied and the pooled analysis did not suggest a statistically significant difference compared with placebo (SMD 0.86; CI 1.11, 2.83; 3 RCTs; I^2 96%), and heterogeneity was high.

One of the studies (Uebelhack et al., 2004) reported on the outcome remission and found a statistically significant effect favoring SJW over placebo (RR 0.13; CI 0.05, 0.36; 1 RCT).

One RCT (Philipp, Kohnen, and Hiller, 1999) in this subgroup also examined quality of life. The study reported a statistically significant positive effect for the mental (SMD 0.46; CI 0.11, 0.81) and the physical component (SMD 0.35; 0.01, 0.70) of the SF-36.

Two RCTs in this subgroup reported serious adverse events; all five occurred in placebo groups. There were statistically significantly fewer neurologic/nervous system adverse events in the SJW groups compared with placebo (OR 0.35; 0.14, 0.82; 2 RCTs). In this subgroup analysis, there was a somewhat lower rate of gastrointestinal/metabolic/nutritional adverse events, but results were not statistically significantly different between study arms (OR 0.77; CI 0.40, 1.48; 3 RCTs).

Severe Depression

We identified no RCT reporting on severe depression comparing SJW and placebo.

One of the included studies (Laakmann, Dienel, and Kieser, 1998) noted that for the more severely depressed patients (HAMD score <23 at baseline), there was a greater reduction in HAMD scores for those taking WS 5572 (12.0 +/− 3.7 points) compared with those taking WS 5573 or placebo (6.6+/− 7.7 points and 7.8+/− 5.4 points), but this comparison was not addressed in any other study.

One RCT (Shelton et al., 2001) described patients' depression severity as moderate or severe. Results did not show statistically significant effects for the number of responders (RR 0.70; CI 0.42, 1.18; 1 RCT), and the depression scale analysis was borderline (SMD 0.28; CI 0.00, 0.56; 1 RCT). The study showed a large effect for the outcome remission, but with wide confidence intervals given the small size of the study (RR 6.73; CI 4.13, 10.94; 1 RCT). The study did not report on quality of life. The study reported more neurologic/nervous system adverse events in the SJW group compared with placebo (OR 2.08; CI 1.09, 4.04; 1 RCT).

Meta-Regression

A meta-regression aiming to identify an association between the depression severity and the size of the treatment effect of SJW compared with placebo did not show statistically significant effects (outcome responders p=0.798; depression scale scores p=0.365; remission p=0.159; number of patients with adverse events p=0.480) in the included studies. Analyses could be performed only for selected outcomes due to the small number of studies in some subgroups. In addition, the large majority of studies was in samples of combined mild and moderate depression; therefore, the effect of severe depression could not be determined.

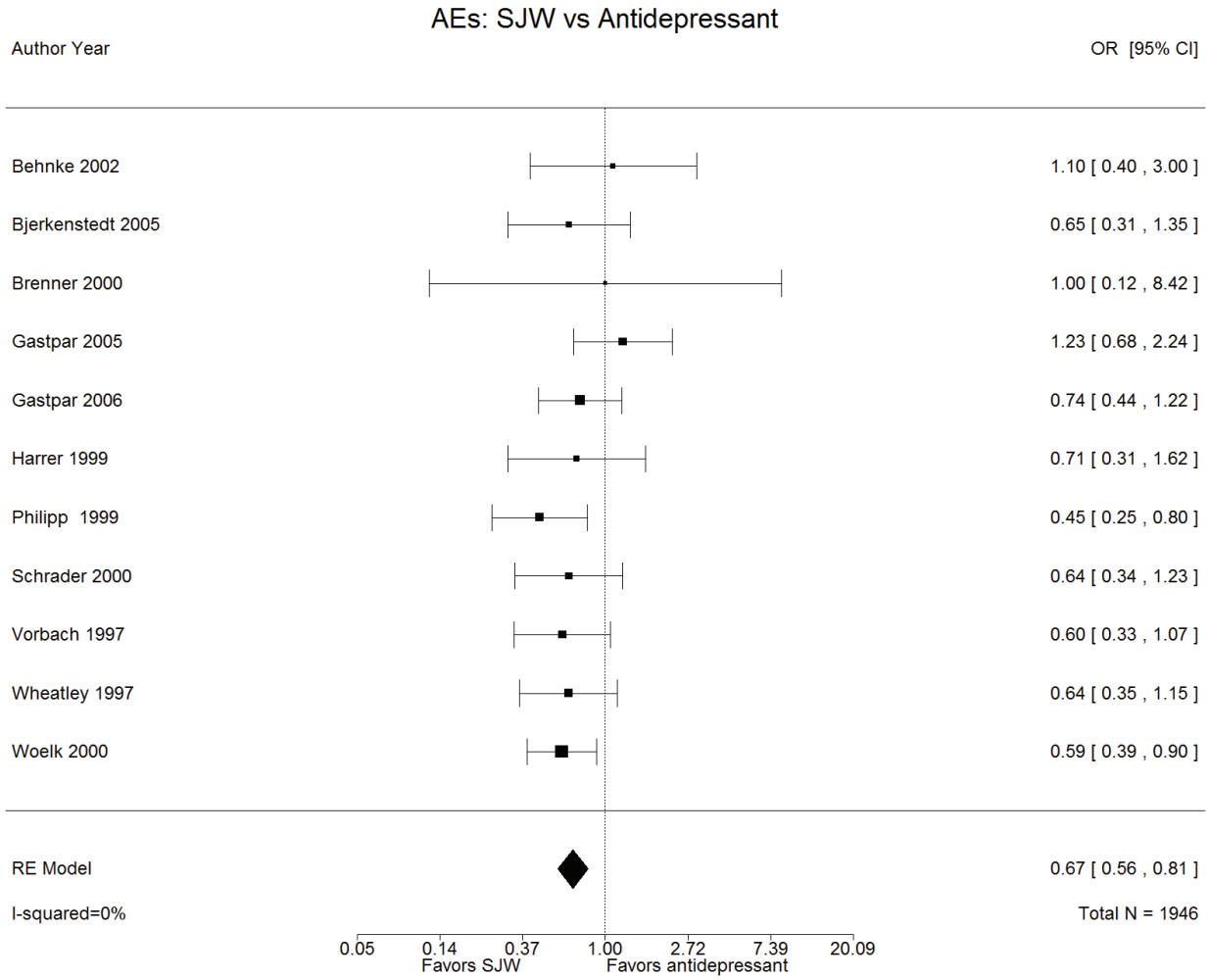
KQ 1d: Are Adverse Events Associated with SJW Comparable to Standard Antidepressant Treatment?

Although most included RCTs addressed the safety of SJW, the assessment rigor varied greatly. In addition, for this comparative analysis, there is likely to be a lack of statistical power because the studies were not powered to show a difference in a rare event. The following subsections report analyses for adverse events assessed in more than one study and counts for rare events that have been associated with SJW (e.g., induction of mania).

Participants with Adverse Events

Eleven RCTs reported the number of participants with adverse events (AE) per intervention group. In all but two individual studies, there was no statistically significant difference between study arms, but the pooled analysis shows a statistically significant effect favoring SJW over antidepressants (OR 0.67; CI 0.56, 0.81; 11 RCTs), as shown in Figure 3.8.

Figure 3.8. St. John's Wort Versus Antidepressants, Number of Participants with Adverse Events



As a sensitivity analysis, we also used the total number of events reported per treatment arm. This analysis indicated the same results for the point estimate, but, as outlined, the CIs are underestimated in this analysis and may be wider than shown, given that the same participant may experience more than one adverse event (OR 0.64; est. CI 0.44, 0.93; 10 RCTs).

Four studies reported serious adverse events, those requiring hospitalization; across studies there were fewer events in the SJW group compared with antidepressants, but the difference was not statistically significant between interventions (OR 0.62; CI 0.05, 5.46; 4 RCTs). Regarding adverse events that have been associated with SJW in the literature, one study (van Gurp et al., 2002) reported that a patient in the SJW group developed an acute manic reaction and was subsequently hospitalized. A patient with hypertensive crisis was reported in one study (Szegedi et al., 2005) comparing WS 5570 and paroxetine, but study authors assessed it as unrelated to the SJW intervention because another cause was evident.

Specific Adverse Events

We also investigated individual adverse events grouped into clinically similar categories.

Fifteen studies reported gastrointestinal/metabolic-nutritional adverse events; the pooled analyses indicated statistically significantly more events in the antidepressant study arms compared with SJW (OR 0.43; CI 0.34, 0.55; 15 RCTs). This effect was essentially unchanged when systematically excluding poor quality studies (OR 0.37; CI 0.28, 0.48; 9 RCTs).

Fifteen studies reported on neurologic/nervous system events; pooled analyses showed statistically significantly more events in the antidepressant study arms compared with SJW (OR 0.29; CI 0.24, 0.36; 15 RCTs). The effect was stable when excluding poor quality studies (OR 0.25; CI 0.20, 0.32; 9 RCTs).

Studies reporting on psychiatric events showed statistically significantly more events in the antidepressant study arms compared with SJW (OR 0.41; CI 0.19, 0.87; 4 RCTs). The effect was not statistically significant when excluding a poor quality study (OR 0.51; CI 0.23, 1.14; 3 RCTs).

Two studies reported on sexual dysfunction events; the pooled analysis showed statistically significantly more events in the antidepressant study arms compared with SJW (OR 0.51; CI 0.30, 0.88; 2 RCTs); both studies were fair/good quality studies.

Studies that reported skin/musculoskeletal events showed a trend for more adverse events in the SJW intervention groups, but the difference was not statistically significant across studies (OR 1.18; CI 0.79, 1.78; 10 RCTs). The same result was seen after excluding poor quality studies (OR 1.17; CI 0.74, 1.84; 6 RCTs). One RCT (Pakseresht et al., 2012) reported photosensitivity for three of 20 patients taking SJW. Of note, two relevant studies excluded participants with known photosensitivity from the trial (Gastpar, Singer, and Zeller, 2005; Gastpar, Singer, and Zeller, 2006).

Five studies reported on cardiovascular events and suggested somewhat more events in antidepressant study arms, but there was no statistically significant difference (OR 0.55; CI 0.26, 1.16; 5 RCTs), even after excluding poor quality studies (OR 0.72; CI 0.32, 1.59; 3 RCTs).

Four studies reported on other organ system (e.g., eye, ear, liver, renal, reproductive) events; pooled analyses showed somewhat fewer events in the SJW interventions compared with antidepressants, but the difference was not statistically significant (OR 0.85; CI 0.52, 1.38; 4 RCTs), even when a poor quality RCT was excluded (OR 0.82; CI 0.50, 1.34; 3 RCTs).

Two studies reported on respiratory/infectious events; pooled analyses showed somewhat more events in the SJW groups, but no statistically significant difference between interventions (OR 1.25; CI 0.70, 2.25; 2 RCTs), even when the poor quality study was excluded (OR 1.14; CI 0.54, 2.46; 1 RCT).

Type of Standard Antidepressant Treatment

In the largest group of antidepressants used in studies included in this review, SSRIs, subgroup results were similar to the main analysis, but the difference in the number of participants with adverse events was not statistically significant (OR 0.81; CI 0.63, 1.04; 7 RCTs). The sensitivity analysis, using the number of total adverse events reported in study arms, showed SJW to be superior (OR 0.65, est. CI 0.57, 0.74; 5 RCTs). There were somewhat fewer serious adverse events in the SJW group compared with antidepressants, but the difference was not statistically significant between interventions (OR 0.62; CI 0.05, 5.46; 3 RCTs).

The SSRI RCTs reporting on gastrointestinal/metabolic-nutritional events indicated statistically significantly more events in the antidepressant study arm compared with SJW (OR 0.40; CI 0.30, 0.53; 9 RCTs). Studies reporting on neurologic/nervous system events also showed statistically significantly more events in the antidepressant study arm (OR 0.62; CI 0.46, 0.84; 9 RCTs). This subgroup included all RCTs that reported on psychiatric events (OR 0.41; CI 0.19, 0.87; 4 RCTs) and those that reported sexual dysfunction events (OR 0.51; CI 0.30, 0.88; 2 RCTs). Trends were also similar to what was seen in the main analysis for skin/musculoskeletal events (OR 1.26; CI 0.82, 1.92; 7 RCTs) and cardiovascular events (OR 0.81; CI 0.29, 2.22; 2 RCTs), and were identical for other organ system (e.g., eye, ear, liver, renal, reproductive) events (OR 0.85; CI 0.52, 1.38; 4 RCTs) and respiratory/infectious events (OR 1.25; CI 0.70, 2.25; 2 RCTs).

For the second subgroup, tricyclic antidepressants, three studies reporting on the outcome number of participants with adverse events showed statistically significantly more events in the antidepressant study arms compared with SJW (OR 0.43; CI 0.25, 0.72; 3 RCTs), and the sensitivity analysis confirmed this result (OR 0.57; est. CI 0.45, 0.71; 4 RCTs). One RCT reported on the absence of serious adverse events in both groups. This subgroup also showed statistically significant results for gastrointestinal/metabolic-nutritional adverse events favoring SJW (OR 0.50; CI 0.29, 0.82; 4 RCTs). RCTs reporting on neurologic/nervous system events also showed statistically significantly more events in the antidepressant study arm in this subgroup (OR 0.13; CI 0.10, 0.19; 4 RCTs). Protective effects of SJW for cardiovascular adverse events were not statistically significant in this subgroup (OR 0.44; CI 0.10, 1.62; 2 RCTs). There was no difference in skin/musculoskeletal events (OR 0.92; CI 0.12, 6.97; 2 RCTs).

Only one RCT compared SJW and psychotherapy (Liu et al., 2010), and the study reported adverse events only for the SJW group (e.g., thirstiness, constipation, nausea, and dizziness); therefore, there was no comparative analysis of adverse events.

Treatment Modality Effects

As outlined, all but one included study tested SJW as monotherapy and did not systematically assess its effect as adjunctive treatment to other antidepressant treatment

interventions. The adjunctive therapy study (Pakseresht et al., 2012) does not contribute to the comparison of SJW versus antidepressants.

To determine whether the reported adverse events were potentially influenced by patients taking antidepressants in both groups (i.e., some SJW patients may continue using already prescribed antidepressants during the research study), we conducted a sensitivity analysis for all trials that explicitly stated that SJW could not use antidepressants. Adverse event results were similar in this subgroup. One additional statistically significant difference between groups emerged in this subgroup: There were statistically significantly more cardiovascular adverse events in the antidepressant arms (OR 0.35; 0.13, 0.87; 4 RCTs). However, the effect for psychiatric events in favor of SJW was not statistically significant in this subgroup (OR 0.44; CI 0.18, 1.07; 2 RCTs).

Extract Effects

We also grouped the existing studies by SJW extract. Four studies reported on the number of participants with adverse events in the LI 160 subgroup; the result was very similar to the main analysis (OR 0.63; CI 0.54, 0.74; 4 RCTs). The sensitivity analysis confirmed this effect (OR 0.54; est. CI 0.35, 0.84; 4 RCTs). We only found more than one study reporting on the same specific adverse events in the LI 160 subgroup. The statistically significant effect for gastrointestinal/metabolic-nutritional adverse events favoring LI 160 over antidepressants was also shown in this subgroup (OR 0.41; CI 0.29, 0.58; 6 RCTs). Across studies, statistically significantly more adverse neurologic/nervous system events were reported for the antidepressant study arm compared with LI 160 (OR 0.49; CI 0.37, 0.66; 6 RCTs). There was also a statically significant effect for cardiovascular events favoring LI 160 because all reported adverse events occurred in the control group (OR 0; CI 0.00, 0.79; 2 RCTs). However, there were statistically significantly more skin/musculoskeletal events in the LI 160 groups compared with antidepressants in this subset (OR 1.79; CI 1.03, 3.16; 6 RCTs).

Depression Severity Effects

We also analyzed whether the severity of the depression makes a difference when comparing SJW and antidepressants in subgroups that contained more than one RCT. However, the available evidence is primarily based on participants with mild and moderate depression; hence, differential effects of antidepressants versus SJW specifically in mild depression cannot be assessed, and only one RCT tested SJW in severe depression.

The effect for neurologic/nervous system adverse events in favor of SJW was also statistically significant in those with moderate depression (OR 0.13; CI 0.06, 0.25; 3 RCTs), as was the effect for other organ system (e.g., eye, ear, liver, renal, reproductive) adverse events, showing more events in the antidepressant group (OR 0.25; CI 0.04, 0.96; 2 RCTs). The studies in this subgroup reporting on gastrointestinal/metabolic/nutritional adverse events showed somewhat more events in the antidepressant groups, but there was no statistically significant

difference (OR 0.61; CI 0.36, 1.00; 3 RCTs). In this subgroup, there was no difference between interventions in reported skin/musculoskeletal adverse events (OR 1.05; CI 0.41, 2.71; 2 RCTs).

In the largest subgroup, studies included participants with mild or moderate depression; the number of participants with adverse events showed similar results to the main analysis, with studies showing fewer patients with adverse events in the SJW intervention groups (OR 0.65; CI 0.56, 0.77; 7 RCTs). Analyzing the total number of events per intervention group in this subgroup, however, showed no statistically significant effect in favor of SJW (OR 0.81; est. CI 0.45, 1.47; 6 RCTs). In this subgroup, we also found the statistically significant effect in favor of SJW for gastrointestinal/metabolic-nutritional events (OR 0.38; CI 0.22, 0.65; 5 RCTs) and the effect for neurologic/nervous system events showing more events in the antidepressant study arm (OR 0.21; CI 0.14, 0.30; 5 RCTs). There were somewhat more skin/musculoskeletal adverse events in the SJW group in this subgroup, but there was no statistically significant effect between interventions (OR 1.48; CI 0.56, 4.09; 3 RCTs). Studies reporting on psychiatric adverse events showed a trend for fewer events in the SJW in this subgroup, but there were no statistically significant differences between interventions (OR 0.35; CI 0.06, 1.53; 2 RCTs).

One RCT in severe depression (Vorbach, Arnoldt, and Hubner, 1997) reported a statistically significant effect in favor of LI 160 compared with imipramine for the number of patients experiencing an adverse event (OR 0.12; CI 0.06, 0.24; 1 RCT). The study reported statistically significantly fewer gastrointestinal/metabolic-nutritional adverse events (OR 0.31; CI 0.08, 0.96; 1 RCT), as well as neurologic/nervous system events (OR 0.14, 0.51; 1 RCT), in the SJW group. The three reported cardiovascular events occurred in the antidepressant groups (OR 0.00; CI 0.00, 2.29; 1 RCT), and there were somewhat fewer skin/musculoskeletal events in the SJW group, but there were no statistically significant differences between interventions (OR 0.47; CI 0.01, 9.22; 1 RCT).

Meta-Regressions for Study-Level Characteristics

A meta-regression did not suggest an association with the number of adverse events reported in the interventions and the severity of depression in included patient samples (p=0.762). However, the majority of existing studies did not differentiate between mild and moderate depression, and only one study tested effects in severe depression.

We found no evidence suggesting publication bias based on the number of participants with adverse events data (Egger test p=0.866, Begg test p=0.773) or across all individual adverse event categories reported across studies (Egger test p=0.441, Begg test p=0.902).

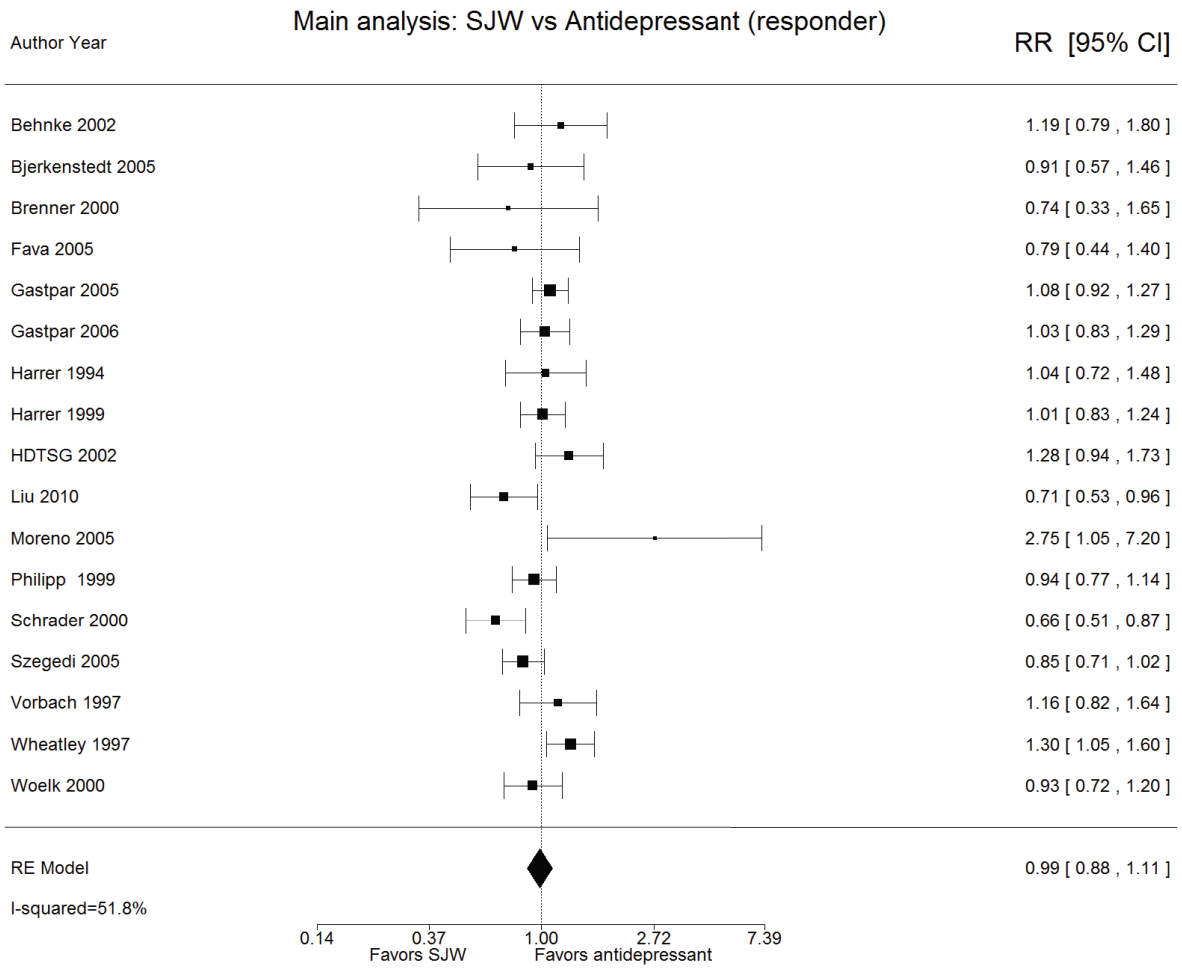
KQ 1e: Is the Efficacy of SJW Comparable to Standard Antidepressant Treatment?

We identified 17 RCTs reporting on the comparative effectiveness of SJW compared with standard antidepressant treatment (i.e., antidepressant medication or psychotherapy). The large majority of trials compared SJW with antidepressant medication, specifically SSRIs.

Depression Treatment Responders

We found 17 RCTs that compared SJW with antidepressants and reported on the number of treatment responders. The median follow-up time was six weeks (range four to 12 weeks). The individual RCTs sometimes favored the antidepressant arm and sometimes the SJW arm, with no clear direction of effects (see Figure 3.9). The pooled analysis did not suggest that effects of SJW and antidepressant medications differ (RR 0.99; CI 0.88, 1.11; 17 RCTs; I^2 53%), although there was some indication of heterogeneity.

Figure 3.9. St. John's Wort Versus Antidepressants, Treatment Responder

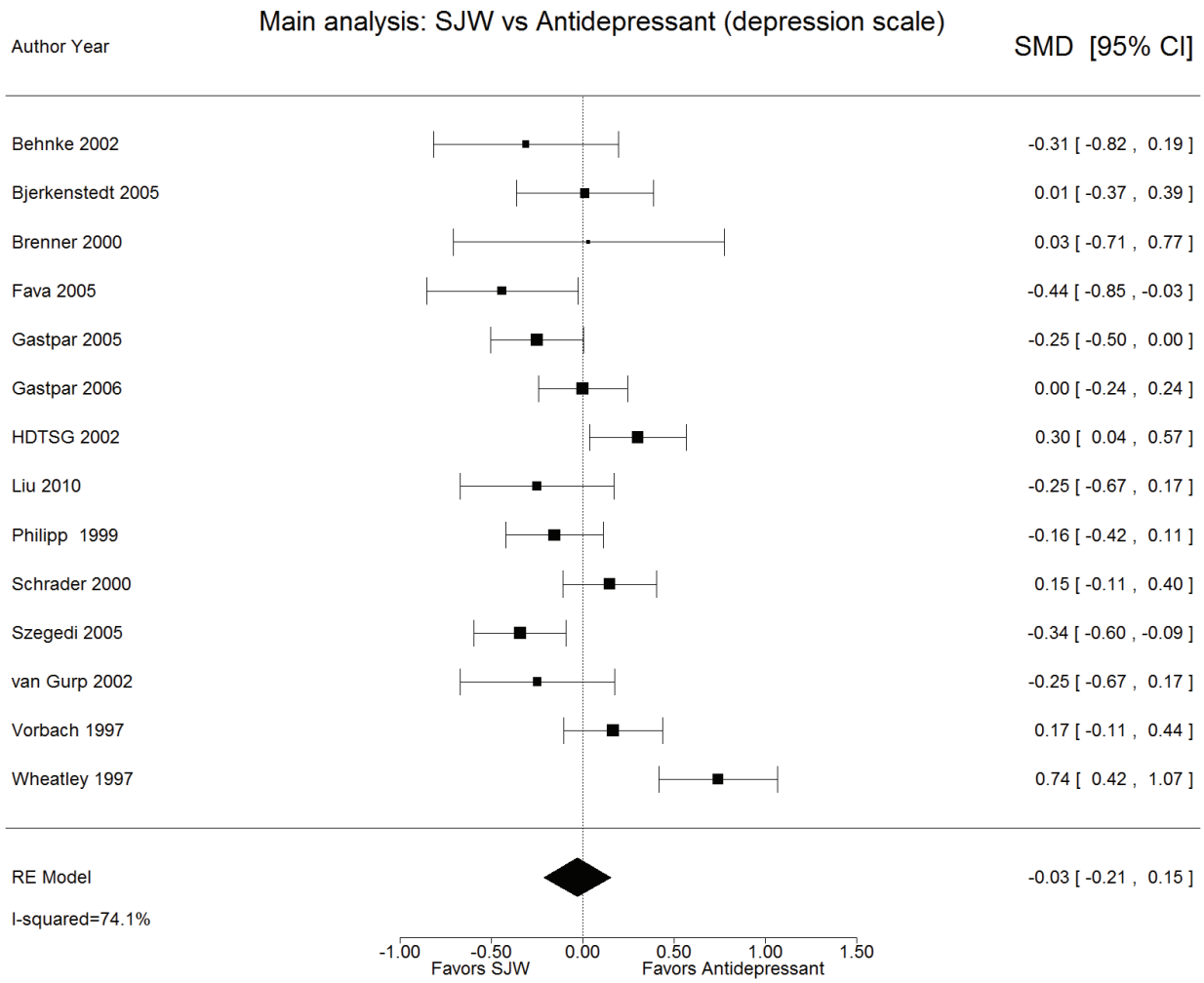


A sensitivity analysis for studies that reported a statistical power analysis showed very similar results (RR 1.02; CI 0.84, 1.25; 5 RCTs; I^2 59%) and also indicated no systematic difference between SJW and antidepressants. A similar result was seen when excluding all poor quality studies (RR 0.98; CI 0.83, 1.16; 11 RCTs; I^2 60%).

Depression Score Standardized Mean Differences

Fourteen RCTs reported on a continuous outcome to measure depressive symptoms. Results were similar to the categorical variable responder; individual studies sometimes favored SJW, and sometimes the antidepressant arms and the pooled analysis did not suggest a different treatment effect between interventions (SMD 0.03; CI 0.15, 0.21; 14 RCTs; I^2 74%). However, there was evidence of heterogeneity between studies, as shown in Figure 3.10.

Figure 3.10. St. John's Wort Versus Antidepressants, Standardized Mean Differences



The median follow-up time was six weeks (range six to 24 weeks). A sensitivity analysis for studies that reported a power analysis and sufficient power to show differences in treatment

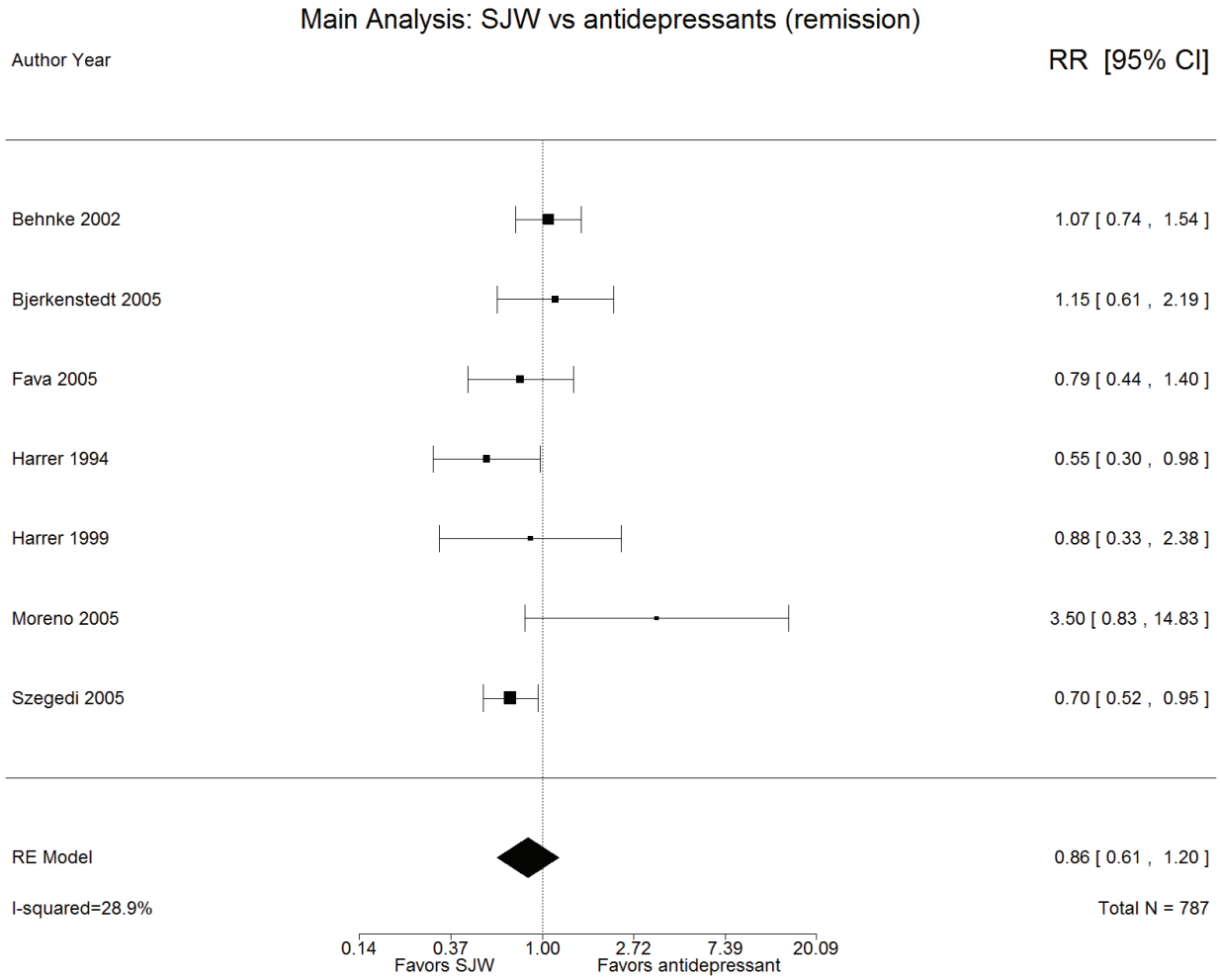
effects between interventions showed an identical point estimate but wider confidence intervals (SMD 0.03; CI 0.75, 0.84; 4 RCTs; I^2 91%), and also indicated no systematic difference between SJW and antidepressants.

Excluding all poor quality RCTs showed no different result (SMD −0.06; CI −0.31, 0.20; 9 RCTs; I^2 78%), and heterogeneity was not reduced.

Depression Remission

We found seven RCTs that compared SJW with antidepressants and measured the number of patients in remission after the intervention (median follow-up six weeks, range four to 12 weeks). Individual studies sometimes favored SJW and sometimes favored antidepressants; the pooled analysis did not show a statistically significant difference between interventions (RR 0.86; CI 0.61, 1.20; 7 RCTs; I^2 19%), as shown in Figure 3.11.

Figure 3.11. St. John's Wort Versus Antidepressants, Remission



There were too few studies reporting power analyses for the outcome remission, but excluding all poor quality studies did also not change the treatment effect in favor of SJW or antidepressants (RR 0.97; CI 0.42, 2.22; 4 RCTs; I^2 44%), and heterogeneity was not reduced in this subsample.

Depression Relapse

One RCT (Gastpar, Singer, and Zeller, 2005) compared SJW with antidepressants and reported on relapse. The study did not show a statistically significant difference between interventions (RR 0.24; CI 0.03, 2.11; 1 RCT).

Quality of Life

One RCT comparing SJW with antidepressants measured quality of life (Philipp, Kohnen, and Hiller, 1999). The study did not show a statistically significant difference between interventions for the mental component of the SF-36 (SMD 0.11; CI −0.15, 0.38; 1 RCT), but results for the physical component were in favor of SJW (SMD 0.35; CI 0.01, 0.70; 1 RCT).

Type of Standard Antidepressant Treatment

We differentiated SSRIs and tricyclic antidepressants from other antidepressant medications to see if there are differences in the comparative effectiveness when comparing SJW and antidepressants.

In the subgroup of SSRIs, we found no difference in effectiveness between SJW and SSRI. The direction of individual studies varied, and the pooled effect estimate was very similar to the main analysis for the outcome responder (RR 0.98; CI 0.83, 1.15; 11 RCTs; I^2 52%). The difference between SJW and SSRIs was not statistically significant for the continuous depression outcome (SMD 0.10; CI −0.08, 0.27; 10 RCTs; I^2 59%). There was also no difference between the interventions for the outcome remission (RR 0.92; CI 0.64, 1.31; 6 RCTs; I^2 0.27%). Heterogeneity in these subgroups was much lower, suggesting that the type of antidepressants may be a source of variance between studies.

In the subgroup of tricyclic antidepressants, we also found no systematic differences in efficacy between the interventions when comparing the number of treatment responders (RR 1.06; CI 0.81, 1.40; 4 RCTs; I^2 52%). Studies reporting differences in depression scales also indicated no difference between interventions (SMD −0.24; CI −1.37, 0.88; 3 RCTs; I^2 90%), but there was high heterogeneity.

Only one RCT reported a comparison with psychotherapy (Liu et al., 2010). The study reported a statistically significant effect in favor of SJW (RR 0.72; CI 0.53, 0.98) for the number of treatment responders but not for the depression scale score (SMD 0.28, CI −0.14, 0.71).

Treatment Modality Effect

As outlined, all but one included study tested SJW as monotherapy and did not systematically assess its effect as adjunctive treatment to other antidepressant treatment interventions (including psychotherapy or antidepressant medication). The adjunctive therapy study (Pakseresht et al., 2012) does not contribute to the comparison of SJW versus antidepressants.

To assess whether the difference between SJW and antidepressant medication was not apparent because a proportion of participants did also use antidepressant medication while taking SJW, we computed a sensitivity analysis for studies explicitly stating that patients on concomitant antidepressant medication were excluded. The pooled analysis for the number of responders was very similar to the main analysis and did not suggest that effects of SJW and antidepressant medications differ (RR 0.98; CI 0.85, 1.13; 14 RCTs; I^2 60%). Furthermore, the sensitivity analyses included the majority of studies contributing to the main analysis. The equivalent pooled analysis for a continuous outcome was very similar to the main analysis and did also not suggest a difference between interventions (SMD 0.02; CI −0.19, 0.23; 12 RCTs; I^2 78%). The pooled analysis for the outcome remission did also not suggest a systematic difference between interventions (RR 0.72; CI 0.42, 1.22; 4 RCTs; I^2 0%), but the absence of heterogeneity in this subset of studies indicated that co-interventions may be associated with the variation in treatment effects across studies.

Extract Effect

We analyzed the existing studies by SJW extract. However, only LI 160 and Ze 117 were used in more than one study reporting on the same outcome.

The subgroup analysis of LI 160 studies did not show statistically significant differences for the number of responders between interventions (RR 1.30; CI 1.05, 160; 6 RCTs; I^2 17%). The subgroup also did not indicate systematic differences for continuous depression measures (SMD −0.12; CI −0.66, 0.42; 5 RCTs; I^2 81%). Finally, there was also no difference for the number of patients in remission (RR 0.78; CI 0.31, 1.95; 3 RCTs; I^2 29%).

One of two studies testing Ze 117 showed a statistically significant effect favoring SJW (RR 0.66; CI 0.51, 0.87; 1 RCT); however, a second study did not replicate this effect (RR 0.93; CI 0.72, 1.20; 1 RCT), and the pooled analysis showed confidence intervals so wide that a pooled effect does not seem to be a good summary of the treatment effect.

Depression Severity Effect

We also analyzed whether the severity of the depression makes a difference when comparing SJW and antidepressants in subgroups that contained more than one RCT. However, the available evidence is primarily based on mild and moderate depression; hence, differential

effects of antidepressants versus SJW specifically in mild depression cannot be assessed and only one study is available for severe depression only.

We did not identify differences between the interventions in the mild and moderate subgroups analyzing the outcome number of responders (RR 1; CI 0.77, 1.30; 8 RCTs; I^2 63%), depression severity (SMD 0.16; CI 0.33, 0.65; 5 RCTs; I^2 76%), or patients in remission (RR 1.12; CI 0.71, 1.76; 4 RCTs; I^2 0%).

In the subgroup of moderate depression severity, there were no differences between interventions for the outcome responder (RR 1.02; CI 0.92, 1.14; 4 RCTs; I^2 0%) and depression severity (SMD 0.13; CI −0.13, 0.45; 3 RCTs; I^2 4%).

One RCT in severe depression (Vorbach, Arnoldt, and Hubner, 1997) reported no statistically significant difference between LI 160 and imipramine for the outcome responder (OR 1.27; CI 0.73, 2.22; 1 RCT) or depression scale scores (SMD −0.17; CI −0.44, 0.11; 1 RCT).

Meta-Regressions for Study-Level Characteristics

Meta-regressions investigating signals suggesting that the comparative effectiveness of SJW varies by extract detected no association (responder outcome p=0.406; SMD p=0.577; remission p=0.236).

Investigating the effect of depression severity did also not suggest a systematic association for the outcome number of responders (p=0.914), SMDs (p=0.503), or remission (p=0.157). However, as discussed, the patient samples in the included studies included primarily mild and moderate depression and did not span the entire range from mild to severe depression.

We did also not identify a systematic association of study quality and treatment effect estimates in comparisons between SJW and antidepressants (responder p=0.378; SMD p=0.105; or remission p=0.654).

An analysis for the type of antidepressant (differentiating SSRIs, tricyclic antidepressants, and other antidepressants) did not suggest a systematic association with the treatment effect estimate (responder outcome p=0.505; SMD p=0.210; remission p=0.236).

There was no evidence suggesting publication bias in the presented dataset for the outcome responder (Egger test p=0.601, Begg test p=0.97), depression scale scores (Egger test 0.506, Begg test p=1), or remission (Egger test 0.247, Begg test 0.381).

Chapter Four: Discussion

Summary of Findings

Overall, the available evidence suggests that SJW extracts are more effective in treating patients with MDD compared with placebo; however, the majority of studies are in patients with mild and moderate MDD using SJW as monotherapy. Observed adverse events were comparable to placebo groups, but not all studies systematically investigated adverse events. We found that the methodological quality of the trials was mostly fair. The presented data are based on trials that used random treatment allocation; presented data are based on ITT analyses where available, and trials were described as double-blind (though some were unclear in how blinding and concealment was ensured). See Table 4.1 for a summary of the evidence.

KQ 1: What Are the Efficacy and Safety of SJW, as an Adjunctive or Monotherapy, for Depressive Symptoms and Quality of Life in Adults with MDD Compared with Placebo or Active Comparator?

In comparing the efficacy of SJW with placebo, we found moderate evidence for SJW over placebo for depression improvement based on the number of participants showing a response to treatment and differences in mean depression scale scores. However, there is low quality evidence of no statistically significant differences in the number of patients in remission. Analyses in placebo-controlled trials demonstrated significant heterogeneity, and subgroup and sensitivity analyses were not able to identify systematic sources of heterogeneity. Effect estimates in individual studies varied, and the unexplained heterogeneity across studies weakens the pooled treatment effect estimate. Treatment effect estimates were similar when studies of poor quality were excluded.

In the studies comparing SJW with placebo, there was moderate evidence that SJW was not more likely to cause adverse events than placebo, overall. However, specific adverse events, such as neurologic/nervous system and organ system (e.g., eye, ear, liver, renal, reproductive) events, were more likely in those taking SJW. The rigor of the adverse event assessments varied; included studies are unlikely to be powered to show statistically significant differences in rare events, and adverse events reported in included RCTs did not systematically address the presence or absence of adverse events highlighted in case reports.

KQ 1a: Is SJW More Effective as Monotherapy Than as Adjunctive Therapy?

We found only one adjunctive study; therefore, the review is not able to determine whether SJW is more effective as monotherapy than as adjunctive therapy. While SJW adjunctive to antidepressant medication is often not recommended due to potential drug interactions (U.S.

Department of Veterans Affairs and U.S. Department of Defense, 2009), we also did not find studies providing SJW adjunctive to psychotherapy. The existing evidence for SJW is based on monotherapy research.

KQ 1b: Is There a Difference in Efficacy, Depending on the Amount and Type of Extract of SJW Used?

We found only one study comparing two different standardized extracts and three studies comparing different dosages, none of which found statistically significant differences between extracts. Nine of the studies did not specify the extract of SJW used. Of those studies that did, the most common extract was LI 160. An indirect comparison of effects of extracts across included studies did not find a difference in results depending on the extract used.

KQ 1c: Is There a Difference in Efficacy, Depending on the Type of MDD (i.e., Mild, Moderate, Severe)?

Analyses did not suggest that the effectiveness or safety of SJW varies by depression severity, but the existing research is primarily based on combined mild and moderate depression patient samples and there is a lack of research studies in severe depression.

The review did not find sufficient evidence to estimate the treatment effect of SJW over placebo for mild depression alone or severe depression alone. Twelve studies tested SJW in mild and moderate depression, and results were similar to the main analysis that included all studies. Three studies provided results for patients with moderate depression compared with placebo and found statistically significant effects compared with placebo in the number of responders and continuous depression outcome, but the pooled analyses did not suggest a statistically significant difference compared with placebo. A meta-regression across the included studies did not show statistically significant effects of an association between the depression severity and the size of the treatment effect of SJW compared with placebo.

KQ 1d: Are Adverse Events Associated with SJW Comparable to Standard Antidepressant Treatment?

All but one of the included studies addressed the safety of SJW, but rigor of assessment varied greatly. In the studies comparing SJW with antidepressant medication, there was moderate evidence that those patients taking antidepressants experienced more adverse events overall. Only one study provided a comparison with psychotherapy. Compared with antidepressants, such as SSRIs, there was low quality evidence showing that SJW is associated with fewer specific adverse events, including gastrointestinal and neurologic adverse events.

There is likely to be a lack of statistical power for the comparative analyses because the studies were not powered to show a difference in these rare events, and included RCTs did not systematically report on the presence or the absence of adverse events that have been linked to SJW, such as increased thyroid stimulating hormone, hypertensive crisis, and induction of mania.

In one included study, one participant had a hypertensive crisis, and in another study, one participant had induction of mania. The included RCTs represent data from about 1,500 participants taking SJW; even assuming that no thyroid events occurred despite the limited adverse event reporting, we can only conclude with 95% confidence that fewer than one person in 400 will experience a thyroid adverse event. All but one included study were monotherapy studies, and only three studies addressed drug interactions (and reported no incidences).

KQ 1e: Is the Efficacy of SJW Comparable to Standard Antidepressant Treatment?

We found no systematic differences in treatment responders (RR 0.99; CI 0.88, 1.11; 17 RCTs, I^2 53%; moderate evidence), depression scale scores (SMD 0.03; CI −0.15, 0.21; 14 RCTs; I^2 74%; moderate evidence), or patients in remission (RR 0.86; CI 0.61, 1.20; 7 RCTs; I^2 29%; low evidence) comparing SJW and antidepressant medication used to treat adults with mild or moderate depression. The effects for the outcome responders and depression scale scores remained stable when limiting to RCTs reporting a power calculation and those with sufficient statistical power to identify an effect; however, the quality of identified studies was poor or fair. Studies reporting on remission were limited in study quality, and the statistical power to detect differences between interventions was unclear. Only one study compared SJW and psychotherapy. There is a lack of data on quality of life. The included studies showed the efficacy of SJW as comparable to antidepressant medication, with SJW being neither inferior nor superior for the treatment of mild or moderate depression.

Table 4.1. Summary of Findings and Quality of Evidence Table

Outcome	Study Design (number of RCTs and participants)	Findings (direction and magnitude of effect)	Study Limitations (study quality; risk of bias)	Inconsistency	Indirectness	Imprecision	GRADE of Evidence for Outcome
KQ 1: Comparison: SJW versus placebo							
Depression, number of treatment responders	18 RCTs, 2,922 participants	RR 0.65 (CI 0.51, 0.84), favors SJW	Majority good or fair quality, effect consistent when excluding poor quality RCTs	Heterogeneity[a]	Direct	Precise	Moderate
Depression scale score	16 RCTs, 2,888 participants	SMD 0.49 (CI 0.23, 0.74), favors SJW	Majority good or fair quality, effect consistent when excluding poor quality RCTs	Heterogeneity[a]	Direct	Precise	Moderate
Depression remission	9 RCTs, 1,419 participants	RR 0.60 (CI 0.22, 1.66), n.s.	Mixed quality, effect consistent when excluding poor quality RCTs but no effect in non-German studies[a]	Heterogeneity, very inconsistent results[a]	Direct	Precise	Low
Depression relapse	1 RCT, 426 participants	RR 1.42 (CI 0.98, 2.06), n.s.	Fair quality	No replication[b]	Direct	Precise	Very Low
Quality of life – mental	2 RCTs, 358 participants	SMD 0.48 (CI 0.24, 0.73), favors SJW	Fair quality	Consistent	Direct	Precise	Low
Quality of life – physical	2 RCTs, 358 participants	SMD 0.28 (CI −1.03, 0.47), n.s.	Fair quality	Inconsistent in small number of RCTs[b]	Direct	Precise	Very Low
Number of patients with adverse events	13 RCTs, 2,600 participants	OR 0.83 (CI 0.62, 1.13), n.s.	Majority good quality, no poor quality studies, RCTs not powered to show effect in rare event[a]	Consistently inconsistent across studies	Direct	Precise	Moderate
Serious adverse events	6 RCTs, 1,358 participants	OR 0.26 (CI 0.04, 1.23), n.s.	Effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Moderate
Gastrointestinal/metabolic-nutritional adverse events	16 RCTs, 2,490 participants	OR 1.06 (CI 0.83, 1.41), no difference	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Neurologic/nervous system adverse events	14 RCTs, 2,404 participants	OR 1.56 (CI 1.08, 3.32), SJW more AEs	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low

Outcome	Study Design (number of RCTs and participants)	Findings (direction and magnitude of effect)	Study Limitations (study quality; risk of bias)	Inconsistency	Indirectness	Imprecision	GRADE of Evidence for Outcome
Skin/musculoskeletal adverse events	10 RCTs, 1,978 participants	OR 1.47 (CI 0.98, 2.21), n.s.	Reporting varied,[a] effect significant when excluding poor quality RCTs[a]	Consistency unclear[a]	Direct	Precise	Very Low
Photosensitivity	4 RCTs, 1,054 participants	OR 1.10 (CI 0.36, 3.56), n.s.	Reporting varied[a]	Consistency unclear[a]	Direct	Precise	Low
Respiratory/infectious adverse events	7 RCTs, 1,081 participants	OR 1.48 (CI 0.95, 2.33), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Other organ system (eye, ear, liver, renal, reproductive) adverse events	5 RCTs, 1,054 participants	OR 1.87 (CI 1.08, 3.32), SJW more AE	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Cardiovascular adverse events	4 RCTs, 759 participants	OR 6.81 (CI 0.92, 304.08), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Imprecise[a]	Low
Psychiatric adverse events	3 RCTs, 608 participants	OR 1.61 (CI 0.34, 10.21), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Imprecise[a]	Low
Sexual dysfunction adverse events	2 RCTs, 428 participants	OR 1.92 (CI 0.94, 4.00), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Imprecise[a]	Very Low

Outcome	Study Design (number of RCTs and participants)	Findings (direction and magnitude of effect)	Study Limitations (study quality; risk of bias)	Inconsistency	Indirectness	Imprecision	GRADE of Evidence for Outcome
KQ 1a: Comparison: SJW as monotherapy versus placebo							
Depression, number of treatment responders	18 RCTs, 2,922 participants	RR 0.65 (CI 0.51, 0.84), favors SJW	Majority good or fair quality, effect consistent when excluding poor quality RCTs	Heterogeneity[a]	Direct	Precise	Moderate
Depression scale score	15 RCTs, 2,848 participants	RR 0.51 (CI 0.24, 0.78), favors SJW	Majority good or fair quality, effect consistent when excluding poor quality RCTs	Heterogeneity[a]	Direct	Precise	Moderate
Depression remission	9 RCTs, 1,419 participants	RR 0.60 (CI 0.22, 1.66), n.s.	Mixed quality but effect consistent when excluding poor quality RCTs	Heterogeneity, very inconsistent results[b]	Direct	Precise	Low
Depression relapse	1 RCT, 426 participants	RR 1.42 (CI 0.98, 2.06), n.s.	Only one fair quality RCT[a]	No replication[b]	Direct	Precise	Very Low
Quality of life – mental	2 RCTs, 358 participants	SMD 0.48 (CI 0.24, 0.73), favors SJW	Only two fair quality studies[b]	Consistent	Direct	Precise	Low
Quality of life – physical	2 RCTs, 358 participants	SMD 0.28 (CI −1.03, 0.47), n.s.	Only two fair quality studies[b]	Inconsistent[a]	Direct	Precise	Very Low
Number of patients with adverse events	13 RCTs, 2,600 participants	OR 0.83 (CI 0.62, 1.13), n.s.	Majority good quality, effect consistent when excluding poor quality RCTs, not powered to show differences in rare event[a]	Consistently inconsistent across individual studies	Direct	Precise	Moderate
Serious adverse events	6 RCTs, 1,358 participants	OR 0.26 (CI 0.04, 1.23), n.s.	Effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Moderate
Gastrointestinal/metabolic-nutritional adverse events	16 RCTs, 2,490 participants	OR 1.06 (CI 0.83, 1.38), no difference	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Neurologic/nervous system adverse events	14 RCTs, 2,404 participants	OR 1.56 (CI 1.08, 3.32), SJW more AEs	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Skin/musculoskeletal adverse events	8 RCTs, 1,368 participants	OR 1.52 (CI 1.00, 2.33), SJW more AEs	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low

Outcome	Study Design (number of RCTs and participants)	Findings (direction and magnitude of effect)	Study Limitations (study quality; risk of bias)	Inconsistency	Indirectness	Imprecision	GRADE of Evidence for Outcome
Respiratory/infectious adverse events	7 RCTs, 1,081 participants	OR 1.48 (CI 0.95, 2.33), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Other organ system (eye, ear, liver, renal, reproductive) adverse events	5 RCTs, 1,054 participants	OR 1.87 (CI 1.08, 3.32), SJW more AEs	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Cardiovascular adverse events	4 RCTs, 759 participants	OR 6.81 (CI 0.92, 304.08), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Imprecise[a]	Very Low
Psychiatric adverse events	3 RCTs, 608 participants	OR 1.61 (CI 0.34, 10.21), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Imprecise[a]	Very Low
Sexual dysfunction adverse events	2 RCTs, 428 participants	OR 1.92 (CI 0.94, 4.00), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs but leaves only 1 RCT[a]	Consistency unclear[a]	Direct	Imprecise[a]	Very Low

Outcome	Study Design (number of RCTs and participants)	Findings (direction and magnitude of effect)	Study Limitations (study quality; risk of bias)	Inconsistency	Indirectness	Imprecision	GRADE of Evidence for Outcome
KQ 1a: Comparison: SJW as adjunctive therapy versus placebo							
Depression scale (continuous outcome)	1 RCT, 40 participants	RR 0.07 (CI −0.55, 0.69), n.s.	Only one, very small, only fair quality RCT[a]	No replication[b]	Direct	Precise	Very Low
Skin/musculoskeletal adverse events	1 RCT, 40 participants	OR N/A (CI 0.43, inf.), n.s.	Only one, very small, only fair quality RCT[a]	No replication[b]	Direct	Imprecise[a]	Very Low
KQ 1b: Comparison: WS 5572 versus WS 5573							
Depression, number of treatment responders	1 RCT, 147 participants	RR 0.80 (CI 0.50, 1.27), n.s.	Good quality	No replication[b]	Direct	Imprecise[a]	Very Low
Depression scale score	1 RCT, 147 participants	SMD 0.40 (CI −0.01, 0.81), n.s.	Good quality	No replication[b]	Direct	Imprecise[a]	Very Low
Number of patients with adverse events	1 RCT, 147 participants	RR 1.04 (CI 0.56, 1.94), no difference	Good quality	No replication[b]	Direct	Imprecise[a]	Very Low
KQ 1b: Comparison: 600 mg versus 1200 mg WS 5570							
Depression, responders	1 RCT, 332 participants	RR 0.88 (CI 0.73, 1.06), n.s.	Only one fair quality RCT[a]	No replication[b]	Direct	Precise	Very Low
Depression scale score	1 RCT, 332 participants	SMD 0.08 (CI −0.18, 0.33), no difference	Only one fair quality RCT[a]	No replication[b]	Direct	Precise	Very Low
Depression remission	1 RCT, 332 participants	RR 1.23 (CI 0.88, 1.72), n.s.	Only one fair quality RCT[a]	No replication[b]	Direct	Imprecise[a]	Very Low
Quality of life – mental	1 RCT, 332 participants	SMD 0.08 (CI −0.17, 0.33), no difference	Only one fair quality RCT[a]	No replication[b]	Direct	Precise	Very Low
Quality of life – physical	1 RCT, 332 participants	SMD 0.14 (CI −0.37, 0.14), no difference	Only one fair quality RCT[a]	No replication[b]	Direct	Precise	Very Low
Number of patients with adverse events	1 RCT, 332 participants	OR 0.99 (CI 0.73, 1.34), no difference	Only one fair quality RCT[a]	No replication[b]	Direct	Precise	Very Low
KQ 1b: Comparison: 0.17 mg, 0.33 mg, 1 mg hypericin							
Depression scale score	1 RCT, 348 participants	3 factor variance analysis, n.s.	Poor quality[a]	No replication[b]	Direct	Imprecise[a]	Very Low
Number of patients with adverse events	1 RCT, 348 participants	Fisher's exact test, n.s.	Poor quality[a]	No replication[b]	Direct	Imprecise[a]	Very Low

Outcome	Study Design (number of RCTs and participants)	Findings (direction and magnitude of effect)	Study Limitations (study quality; risk of bias)	Inconsistency	Indirectness	Imprecision	GRADE of Evidence for Outcome
KQ 1c: Effect of depression severity							
Depression, responders	18 RCTs, 2,922 participants	Meta-regression did not suggest differences between patient subgroups (p=0.798)	Majority good or fair quality	Majority mild and moderate, no severe depression data[b]	Indirect[a]	Precise	Very Low
Depression scale score	16 RCTs, 2,888 participants	Meta-regression did not suggest differences between patient subgroups (p=0.365)	Majority good or fair quality	Majority mild and moderate, no severe depression data[b]	Indirect[a]	Precise	Very Low
Depression remission	9 RCTs, 1,507 participants	Meta-regression did not suggest differences between patient subgroups (p=0.159)	Mixed quality	Majority mild and moderate, no severe depression data[b]	Indirect[a]	Precise	Very Low
Number of patients with adverse events	13 RCTs, 2,600 participants	Meta-regression did not suggest differences between patient subgroups (p=0.480)	Majority good quality, effect consistent when excluding poor quality RCTs	Majority mild and moderate, no severe depression data[b]	Indirect[a]	Precise	Very Low

Outcome	Study Design (number of RCTs and participants)	Findings (direction and magnitude of effect)	Study Limitations (study quality; risk of bias)	Inconsistency	Indirectness	Imprecision	GRADE of Evidence for Outcome
KQ 1d: Comparison: SJW versus antidepressant							
Number of patients with adverse events	11 RCTs, 1,946 participants	OR 0.67 (CI 0.56, 0.81), favoring SJW	Majority good quality, effect consistent when excluding poor quality RCTs	Consistent but not all studies reported on outcome[a]	Direct	Precise	Moderate
Serious adverse events	4 RCTs, 805 participants	OR 0.62 (CI 0.05, 5.46) n.s.	Effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Imprecise[a]	Low
Gastrointestinal/metabolic-nutritional adverse events	15 RCTs, 2,491 participants	OR 0.43 (CI 0.34, 0.55) favoring SJW	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Neurologic/nervous system adverse events	15 RCTs, 2,492 participants	OR 0.29 (CI 0.24, 0.36), favoring SJW	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Skin/musculoskeletal adverse events	10 RCTs, 1,587 participants	OR 1.18 (CI 0.79, 1.78), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Respiratory/infectious adverse events	2 RCTs, 352 participants	OR 1.25 (CI 0.70, 2.25), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs but only 1 RCT left[a]	Consistency unclear[a]	Direct	Precise	Very Low
Other organ system (eye, ear, liver, renal, reproductive) adverse events	4 RCTs, 761 participants	OR 0.85 (CI 0.52, 1.38), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Cardiovascular adverse events	5 RCTs, 750 participants	OR 0.55 (CI 0.26, 1.16), n.s.	Reporting varied,[a] effect consistent when excluding poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low
Psychiatric adverse events	4 RCTs, 552 participants	OR 0.41 (CI 0.19, 0.87), favoring SJW	Reporting varied,[a] effect not significant when excluding poor quality RCTs[a]	Consistency unclear[a]	Direct	Precise	Very Low
Sexual dysfunction adverse events	2 RCTs, 301 participants	OR 0.51 (CI 0.30, 0.88), favoring SJW	Reporting varied, [a] effect consistent because no poor quality RCTs	Consistency unclear[a]	Direct	Precise	Low

Outcome	Study Design (number of RCTs and participants)	Findings (direction and magnitude of effect)	Study Limitations (study quality; risk of bias)	Inconsistency	Indirectness	Imprecision	GRADE of Evidence for Outcome
KQ 1e: Comparison: SJW versus antidepressants							
Depression, number of treatment responders	17 RCTs, 2,776 participants	RR 0.99 (CI 0.88, 1.11), no difference	Majority fair or poor quality,[a] effect consistent when limited to RCTs with power analysis or excluding poor quality RCTs	Consistently inconsistent across individual studies	Direct	Precise	Moderate
Depression scale score	14 RCTs, 2,248 participants	SMD 0.03 (CI −0.15, 0.21), no difference	Majority fair or poor quality,[a] effect consistent when limited to RCTs with power analysis or excluding poor quality RCTs	Consistently inconsistent across individual studies	Direct	Precise	Moderate
Depression remission	7 RCTs, 787 participants	RR 0.86 (CI 0.61, 1.20), n.s.	Fair and poor quality,[a] effect consistent when excluding poor quality RCTs, not enough powered studies[a]	Consistently inconsistent across individual studies	Direct	Precise	Low
Depression relapse	1 RCT, 241 participants	RR 0.24 (CI 0.03, 2.11), n.s.	Poor quality[a]	No replication[b]	Direct	Precise	Very Low
Quality of life – mental	1 RCT, 216 participants	SMD −0.11 (CI −0.15, 0.38), n.s.	Only one fair quality RCT[a]	No replication[b]	Direct	Precise	Very Low
Quality of life – physical	1 RCT, 153 participants	SMD 0.35 (CI 0.01, 0.70), favors SJW	Only one fair quality RCT[a]	No replication[b]	Direct	Precise	Very Low

NOTE: SJW = St. John's wort; AE = adverse events; n.s. = not statistically significantly different.
[a] Downgrade quality of evidence by one.
[b] Downgrade quality of evidence by two.

Other Reviews in this Area

The results of this review are comparable to the conclusions of a previous review of SJW for major depression by Linde, Berner, and Kriston in 2008, which found that SJW extracts are superior to placebo for MDD, are similarly effective as standard antidepressants, and have fewer side effects than standard antidepressants. Our review includes 28 of the 29 studies from that review (Behnke et al., 2002; Bjerkenstedt et al., 2005; Brenner et al., 2000; Fava et al., 2005; Gastpar, Singer, and Zeller, 2005; Gastpar, Singer, and Zeller, 2006; Hänsgen, Vesper, and Ploch, 1994; Harrer, Hübner, and Podzuweit, 1994; Harrer et al., 1999; HDTSG, 2002; Kalb, Trautmann-Sponsel, and Kieser, 2001; Kasper et al., 2006; Laakmann, Dienel, and Kieser, 1998; Lecrubier et al., 2002; Montgomery, Hübner, and Grigoleit, 2000; Moreno et al., 2005; Philipp, Kohnen, and Hiller, 1999; Schrader, Meier, and Brattström, 1998; Schrader, 2000; Shelton et al., 2001; Szegedi et al., 2005; Uebelhack et al., 2004; van Gurp et al., 2002; Volz, Eberhardt, and Grill, 2000; Vorbach, Arnoldt, and Hubner, 1997; Wheatley, 1997; Witte et al., 1995; Woelk, 2000). One of the trials could not be retrieved (Bracher, 2001). This review added an additional seven studies (Bernhardt, Liske, and Ebeling, 1993; Kasper et al., 2008; Lenoir, Degenring, and Saller, 1999; Liu et al., 2010; Mannel et al., 2010; Pakseresht et al., 2012; Rahman et al., 2008).

A more recent systematic review of pharmacological treatments for depressive disorders in primary care included ten of the studies above and four that did not meet our criteria for inclusion in this review (Linde, 2015). The findings were consistent with the previous review, in that hypericum extracts showed similar efficacy and better acceptability than antidepressants and are effective for treating acute depression, though effects when compared with placebo were modest.

Strengths and Limitations

This review has several strengths: an *a priori* research design, duplicate study selection and data abstraction of study information, a comprehensive search of electronic databases without language restriction, risk of bias assessments, and comprehensive quality of evidence assessments used to formulate review conclusions. However, some limitations are worth noting. First, we did not contact individual study authors; results reported in the review are based on published data. Some of the studies were of poor quality, primarily due to lack of ITT analysis or poor follow-up. The depression improvements associated with SJW were seen in the analyses of the number of responders, as well as mean depression scale scores; however, both treatment effect estimates showed heterogeneity. A large number of subgroup and sensitivity analyses did not identify systematic sources of differences between studies, and heterogeneity remains as a limitation of the SJW evidence. Adverse event evidence is limited because the rigor of adverse event assessments varied greatly, comparative analyses were potentially limited due to the lack

of statistical power to show differences in individual rare events, and RCTs assessed only a limited range of potential adverse events.

The identified studies tested SJW as monotherapy; hence, we could not determine whether SJW is more effective as monotherapy than as adjunctive therapy. Due to the lack of head-to-head trials of different extracts and dosage and the absence of extract effects in indirect comparisons, it is not possible to say which extract or dose is best. Most of the studies focused on mild and moderate depression, and clinicians need to be aware that results of the review may not extrapolate to include all patients with MDD.

Implications for Future Research and Practice

Our conclusions are mostly in line with other reviews in this area: SJW is an effective treatment for mild and moderate MDD, with fewer adverse effects than standard antidepressants. However, future research in this area should include more head-to-head trials between specific extracts and dosage of SJW. While potential risks of drug interactions hinder research of SJW as an adjunctive treatment, research studies on SJW concomitant to psychotherapy are also missing. As quality of life is greatly affected by MDD, it would be good to see more studies of depression treatment include this measure.

As for clinical practice, despite the positive findings of this report, concerns remain. The U.K. Guidelines for Depression in Adults, for example, advise not to prescribe SJW because of uncertainty about appropriate doses, persistence of effect, variation in the nature of preparations, and potential serious interactions with other drugs (including oral contraceptives, anticoagulants, and anticonvulsants). Our review was unable to dismiss these concerns due to lack of trials addressing them. However, the extract LI 160 has been reviewed in a number of studies, and several have reported successful dosing schemes. While reports of rare adverse events—such as hypertensive crisis and induction of mania—cannot be dismissed based on RCT data, it is noteworthy that SJW appears to have fewer adverse events than antidepressant medication in comparative analyses.

Appendix A: Search Strategy

PubMed

Time Period Covered:
1/1/2007-11/24/2014

Search Strategy:
"Hypericum"[Mesh] OR john's wort OR johns wort OR hyperic* OR johanniskraut
AND
"Depressive Disorder"[Mesh] OR "Depression"[Mesh] OR depress*[tiab] OR unipolar OR mood disorder* OR mood disturbance* OR affective disorder*

Number of Results: 264

==========

CINAHL (Cumulative Index to Nursing and Allied Health Literature)

Time Period Covered:
1/1/2007-11/24/2014

Search Strategy:
john's wort OR johns wort OR hyperic* OR johanniskraut
AND
depress* OR unipolar OR mood disorder* OR mood disturbance* OR affective disorder*

Number of Results: 128

==========

PsycINFO

Time Period Covered:
1/1/2007-11/24/2014

Search Strategy:
john's wort OR johns wort OR hyperic* OR johanniskraut
AND
depress* OR unipolar OR mood disorder* OR mood disturbance* OR affective disorder*

Number of Results: 82

==========

Cochrane Central Register of Controlled Trials (CENTRAL)

Time Period Covered:
1/1/2007-11/24/2014

Search Strategy:
john's wort or johns wort or hyperic* or johanniskraut:ti,ab,kw
AND
depress* or unipolar or mood disorder* or mood disturbance* or affective disorder*:ti,ab,kw

Number of Results: 50

==========

Embase

Time Period Covered:
1/1/2007-11/24/2014

Search Strategy:
john* NEAR/2 wort OR (johns AND wort) OR hyperic* OR johanniskraut
AND
depress* OR unipolar OR (('mood'/exp OR mood) AND disorder*) OR (('mood'/exp OR mood) AND disturbance*) OR (affective AND disorder*)
AND
Human

Number of Results: 70

==========

AMED (Allied and Complementary Medicine Database)

Time Period Covered:
1/1/2007-11/24/2014

Search Strategy:
john's wort or johns wort or hyperic* or johanniskraut
and
(depress* or unipolar or mood disorder* or mood disturbance* or affective disorder*).af.

Number of Results: 22

==========

MANTIS (Manual, Alternative, and Natural Therapy Index System)

Time Period Covered:
1/1/2007-11/24/2014

Search Strategy:
john's wort or johns wort or hyperic* or johanniskraut
and
(depress* or unipolar or mood disorder* or mood disturbance* or affective disorder*).af.

Number of Results: 45

==========

Web of Science Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH

Time Period Covered:
1/1/2007-1/19/2015

Search Strategy:
TOPIC: (john's wort or johns wort or hyperic* or johanniskraut)
AND
TOPIC: (depress* or unipolar or mood disorder* or mood disturbance* or affective disorder*)

Number of Results: 404

==========

ICTRP (International Clinical Trials Registry Platform)

Time Period Covered:
1/1/2007-1/19/2015

Search Strategy:
john's wort or johns wort or hyperic* or johanniskraut

Number of Results: 3

Total of All Results After Removing Duplicates and Animal-Only Studies: 555

Appendix B: Depression Scale Standard Cut-Points

Scale	Cut-Off Point
Beck Depression Inventory-I	Cut-off for Clinical Diagnosed with Depression: 0–9 = minimal/no depression 10–18 = mild/moderate depression 19–29 = moderate/severe depression 30–63 = severe depression
Beck Depression Inventory-II	0–13 = minimal 14–19 = mild (13-14*= mild) 20–28 = moderate 29–63 = severe
Center for Epidemiologic Studies Depression (CES-D)	CES-D 20: 16 = “significant” or “mild” depressive symptomatology CES-D 10: 11 = recommended as cut-off (equivalent to experiencing 6 symptoms for most of the previous week or a majority of symptoms on 1 or 2 days)
Clinical Diagnosis/Meets DSM Criteria/Major Depression Inventory (MDI)	26 = moderate-severe depression 0–19 = no depression 20–24 = mild depression 25–29 = moderate depression 30–50 = severe depression
Depression Anxiety Stress Scale (DASS)-21 Depression Scale	0–4 = normal 5–6 = mild 7–10 = moderate 11-13 = severe 14+ = extremely severe 12 = recommended cut-point
Depression-Arkansas Scale (D-ARK)	26–37= mild 38–57 = moderate
General Health Questionnaire (GHQ)	4 = usual cut-point
Geriatric Depression Scale (GDS)	GDS-5: $\geq$ 2 = cut-point GDS-15: 5–9 = mild 10–15 = moderate to severe Cut-off scores for GDS-15 Among Special Populations: Cognitive impairment = 8 Dementia = 11 Parkinson’s Disease = 10–11 (but some variation here) Stroke = 11–12 (minor depressed) Post Stroke = 6–7 Elderly home care = 5 GDS Long Form (30 items) 11–20 = mild 21–30 = moderate to severe

Scale	Cut-Off Point
Hamilton Rating Scale for Depression (HAMD)	0–6 = no depression 7–17 = mild depression 18–24 = moderate depression 24+ = severe depression
Hospital Anxiety and Depression Scale	0–7 = no depression 8–10 = “possible case” 11–21 = “probable case” Optimal cut-off point = ≥ 8 for the identification of suspicious cases and ≥ 11 for safe cases on both subscales
Medical Outcomes Study Depression Screen (MOS-D)	0.06 = usual cut-point
Minnesota Multiphasic Personality Inventory (MMPI) Depression Scale	T score of 70 used for MMPI T score of 65 used for MMPI-2
Montgomery-Åsberg Depression Rating Scale (MADRS)	7–19 = mild 20–34 = moderate 35–60 = severe
Montgomery-Åsberg Depression Rating Scale (MADRS)-S	13–19 = mild 20+ = moderate to severe
Patient Health Questionnaire (PHQ)-9	5 = mild 10 = moderate 15 = severe *10 cited as the optimal cut-off point
Primary Care Evaluation of Mental Disorders (PRIME-MD)	1 = usual cut-point
Symptom Checklist (SCL)-20	≥ 1.75 as a cutoff for major depression
SCL-CD6	≥ 17 is indicative of MDD
Symptom Driven Diagnostic System-Primary Care (SDDS-PC)	2 = usual cut-point
Zung Self Assessment Depression Scale (SDS)	50 = mild 60 = moderate 70 = severe
Alasker scale	N/A
Brief Symptom Inventory (BSI)	N/A
Institute for Personality and Ability Testing Depression Scale (IPAT)	N/A
Patient-Reported Outcomes Measurement Information System (PROMIS) Depression	N/A
SCL-90	N/A

N/A = not applicable.

Appendix C: Excluded Full-Text Articles

Reason Excluded: Background

Bitran, S., A. H. Farabaugh, V. E. Ameral, R. A. LaRocca, A. J. Clain, M. Fava, and D. Mischoulon, "Do Early Changes in the HAM-D-17 Anxiety/Somatization Factor Items Affect the Treatment Outcome Among Depressed Outpatients? Comparison of Two Controlled Trials of St John's Wort (Hypericum Perforatum) Versus a SSRI," *International Clinical Psychopharmacology*, Vol. 26, No. 4, July 2011, pp. 206–212. PMID: 21278577

Carpenter, D. J., "St John's Wort and S-Adenosyl Methionine as "Natural" Alternative to Conventional Antidepressants in the Era of the Suicidality Boxed Warning: What Is the Evidence for Clinically Relevent Benefit?" *Alternative Medicine Review*, Vol. 16, No. 1, 2011, pp. 17–39. PMID: 2011304835

Carpenter, D. J., "St. John's Wort and S-Adenosyl Methionine as 'Natural' Alternatives to Conventional Antidepressants in the Era of the Suicidality Boxed Warning: What Is the Evidence for Clinically Relevant Benefit?" *Alternative Medicine Review*, Vol. 16, No. 1, March 2011, pp. 17–39. PMID: 21438644

Dhingra, S., and M. Parle, "Herbal Remedies and Nutritional Supplements in the Treatment of Depression: A Review," *Klinik Psikofarmakoloji Bülteni/Bulletin of Clinical Psychopharmacology*, Vol. 22, No. 3, September 2012, pp. 286–292. PMID: 2012-29669-014

Ernst, E., "Review: St John's Wort Superior to Placebo and Similar to Antidepressants for Major Depression but with Fewer Side Effects," *Evidence-Based Mental Health*, Vol. 12, No. 3, August 2009, p. 78. PMID: 19633246

Freeman, M. P., "Complementary and Alternative Medicine for Perinatal Depression," *Journal of Affective Disorders*, Vol. 112, No. 1–3, January 2009, pp. 1–10. PMID: 18692251

Freeman, M. P., M. Fava, J. Lake, M. H. Trivedi, K. L. Wisner, and D. Mischoulon, "Complementary and Alternative Medicine in Major Depressive Disorder: The American Psychiatric Association Task Force Report," *Journal of Clinical Psychiatry*, Vol. 71, No. 6, June 2010, pp. 669–681. PMID: 20573326

Gahlsdorf, T., R. Krause, M. W. Beal, T. Gahlsdorf, R. Krause, and M. Beal, "Efficacy of St. John's Wort for Treating Mild to Moderate Depression," *Complementary Health Practice Review*, October 2007.

Grobler, A. C., G. Matthews, and G. Molenberghs, "The Impact of Missing Data on Clinical Trials: A Re-Analysis of a Placebo Controlled Trial of Hypericum Perforatum (St Johns

Wort) and Sertraline in Major Depressive Disorder," *Psychopharmacology* (Berl), Vol. 231, No. 9, May 2014, pp. 1987–1999. PMID: 24232445

Hänsgen, KD., "Antidepressive Wirksamkeit Eines Hochdosierten Hyepricum-Extraktes," *Muenchener Medizinesche Wochenschrift*, Vol. 138, No. 3, 1996, pp. 29–33.

Kasper, S., I. G. Anghelescu, A. Szegedi, A. Dienel, and M. Kieser, "Placebo Controlled Continuation Treatment with Hypericum Extract WS 5570 After Recovery from a Mild or Moderate Depressive Episode," *Wiener Medizinesche Wochenschrift*, Vol. 157, No. 13–14, 2007, pp. 362–366. PMID: 17704988

Kasper, S., F. Caraci, B. Forti, F. Drago, and E. Aguglia, "Efficacy and Tolerability of Hypericum Extract for the Treatment of Mild to Moderate Depression," *European Neuropsychopharmacology*, Vol. 20, No. 11, November 2010, pp. 747–765. PMID: 20708905

Kasper, S., M. Gastpar, H. J. Moller, W. E. Muller, H. P. Volz, A. Dienel, and M. Kieser, "Better Tolerability of St. John's Wort Extract WS 5570 Compared to Treatment with SSRIs: A Reanalysis of Data from Controlled Clinical Trials in Acute Major Depression," *International Clinical Psychopharmacology*, Vol. 25, No. 4, July 2010, pp. 204–213. PMID: 20568656

Kasper, S., M. Gastpar, W. E. Muller, H. P. Volz, A. Dienel, M. Kieser, and H. J. Moller, "Efficacy of St. John's Wort Extract WS (R) 5570 in Acute Treatment of Mild Depression," *European Archives of Psychiatry and Clinical Neuroscience*, Vol. 258, No. 1, December 2007, pp. 59–63.

Kasper, S., M. Gastpar, W. E. Muller, H. P. Volz, A. Dienel, and H. J. Moller, "Treatment of Depression with St. John's Wort Extract WS (R) 5570. No Influence on Patients' Weight," *Psychopharmakotherapie*, Vol. 15, No. 1, January–February 2008. PMID: WOS:000252988000003

Laakmann, G., C. Schule, T. Baghai, and M. Kieser, "St. John's Wort in Mild to Moderate Depression: The Relevance of Hyperforin for the Clinical Efficacy," Pharmacopsychiatry, Vol. 31, Suppl. 1, June 1998, pp. 54–59. PMID: 9684948

Linde, K., M. M. Berner, and L. Kriston, "St John's Wort for Major Depression," *The Cochrane Database of Systematic Reviews*, No. 4, 2008, p. CD000448. PMID: 18843608

Luberto, C. M., C. White, R. W. Sears, and S. Cotton, "Integrative Medicine for Treating Depression: An Update on the Latest Evidence," *Current Psychiatry Reports*, Vol. 15, No. 9, September 2013, p. 391. PMID: 23943471

Nahas, R., and O. Sheikh, "Complementary and Alternative Medicine for the Treatment of Major Depressive Disorder," *Canadian Family Physician*, Vol. 57, No. 6, June 2011, pp. 659–663. PMID: 21673208

Papakostas, G. I., C. M. Crawford, M. J. Scalia, and M. Fava, "Timing of Clinical Improvement and Symptom Resolution in the Treatment of Major Depressive Disorder. A Replication of Findings with the Use of a Double-Blind, Placebo-Controlled Trial of Hypericum Perforatum Versus Fluoxetine," *Neuropsychobiology*, Vol. 56, No. 2–3, 2007, pp. 132–137. PMID: 18259086

Purgato, M., D. Papola, C. Gastaldon, C. Trespidi, L. R. Magni, C. Rizzo, T. A. Furukawa, N. Watanabe, A. Cipriani, and C. Barbui, "Paroxetine Versus Other Anti-Depressive Agents for Depression," *The Cochrane Database of Systematic Reviews*, Vol. 4, 2014, p. Cd006531. PMID: 24696195

Rahimi, R., S. Nikfar, and M. Abdollahi, "Efficacy and Tolerability of Hypericum Perforatum in Major Depressive Disorder in Comparison with Selective Serotonin Reuptake Inhibitors: A Meta-Analysis," *Progress in Neuropsychopharmacology and Biologic Psychiatry*, Vol. 33, No. 1, February 1, 2009, pp. 118–127. PMID: 19028540

Sarris, J., M. Fava, I. Schweitzer, and D. Mischoulon, "St John's Wort (Hypericum Perforatum) Versus Sertraline and Placebo in Major Depressive Disorder: Continuation Data from a 26-Week RCT," *Pharmacopsychiatry*, Vol. 45, No. 7, November 2012, pp. 275–278. PMID: 22592504

Sarris, J., A. A. Nierenberg, I. Schweitzer, J. E. Alpert, J. F. Rosenbaum, N. Iovieno, J. Covino, M. Fava, and D. Mischoulon, "Conditional Probability of Response or Nonresponse of Placebo Compared with Antidepressants or St John's Wort in Major Depressive Disorder," *Journal of Clinical Psychopharmacology*, Vol. 33, No. 6, December 2013, pp. 827–830. PMID: 24091858

Singer, A., M. Schmidt, W. Hauke, and K. Stade, "Duration of Response After Treatment of Mild to Moderate Depression with Hypericum Extract STW 3-VI, Citalopram and Placebo: A Reanalysis of Data from a Controlled Clinical Trial," *Phytomedicine*, Vol. 18, No. 8–9, June 15, 2011, pp. 739–742. PMID: 21514125

Steigerwald Arzneimittelwerk GmbH, "Double Blind, Randomized, Placebo-Controlled, Mulitcentric Clinical Trial to Evaluate the Relapse Prevention of a Hypericum Extract in Outdoor Patients with Moderate Depressive Episodes (Major Depression)," 2008. PMID: CN-00974375

Undurraga, J., and R. J. Baldessarini, "Randomized, Placebo-Controlled Trials of Antidepressants for Acute Major Depression: Thirty-Year Meta-Analytic Review,"

Neuropsychopharmacology, Vol. 37, No. 4, March 2012, pp. 851–864. PMID: WOS:000300580100001

Varteresian, T., and H. Lavretsky, "Natural Products and Supplements for Geriatric Depression and Cognitive Disorders: An Evaluation of the Research," *Current Psychiatry Reports*, Vol. 16, No. 8, August 2014, p. 456. PMID: 24912606

Reason Excluded: Excluded Design

Borrelli, F., and A. A. Izzo, "Herb-Drug Interactions with St John's Wort (Hypericum Perforatum): An Update on Clinical Observations," *The AAPS Journal*, Vol. 11, No. 4, December 2009, pp. 710–727. PMID: 19859815

Briese, V., U. Stammwitz, M. Friede, and H. H. Henneicke-von Zepelin, "Black Cohosh With or Without St. John's Wort for Symptom-Specific Climacteric Treatment—Results of a Large-Scale, Controlled, Observational Study," *Maturitas*, Vol. 57, No. 4, August 20, 2007, pp. 405–414. PMID: 17590291

Devrim, O., Y. Ozturk, and U. D. Ozkay, "A Natural Antidepressant: Hypericum Perforatum L.: Review," *Turkiye Klinikleri Tip Bilimleri Dergisi*, Vol. 29, No. 3, June 2009, pp. 708–715. PMID: WOS:000268007400021

Jurcic, J., J. A. Pereira, and D. Kavanaugh, "St John's Wort Versus Paroxetine for Depression," *Canadian Family Physician*, Vol. 53, No. 9, September 2007, pp. 1511–1513. PMID: 17872882

Lake, J., "Natural Products Used to Treat Depressed Mood as Monotherapies and Adjuvants to Antidepressants: A Review of the Evidence," *Psychiatric Times*, Vol. 26, No. 10, 2009, pp. 51–56. PMID: 2010460947

Lecrubier, Y., and P. Bech, "The Ham D-6 Is More Homogenous and as Sensitive as the Ham D-17," *European Psychiatry*, Vol. 22, No. 4, May 2007, pp. 252–255. PMID: WOS:000247605800008

Meeks, T. W., J. L. Wetherell, M. R. Irwin, L. S. Redwine, and D. V. Jeste, "Complementary and Alternative Treatments for Late-Life Depression, Anxiety, and Sleep Disturbance: A Review of Randomized Controlled Trials," *Journal of Clinical Psychiatry*, Vol. 68, No. 10, October 2007, pp. 1461–1471. PMID: WOS:000250620700001

Moss, A. S., D. A. Monti, J. D. Amsterdam, and A. B. Newberg, "Complementary and Alternative Medicine Therapies in Mood Disorders," *Expert Review of Neurotherapeutics*, Vol. 11, No. 7, July 2011, pp. 1049–1056. PMID: WOS:000293853900018

Muller, J., C. Kolb, O. Kelber, and D. Weiser, "High-Dose St. John's Wort Extract STW 3-VI as a Daily Single-Dose Is Safe and Effective in the Teatment of Mild to Moderate Depression in

Different Age-Groups—Results of a Re-Evaluation," *Planta Medica*, Vol. 76, No. 12, August 2010, pp. 1357–1358. PMID: WOS:000282066500760

Pirotta, M., K. Densley, K. Forsdike, M. Carter, and J. Gunn, "St John's Wort Use in Australian General Practice Patients with Depressive Symptoms: Their Characteristics and Use of Other Health Services," *Journal of the Australian Traditional Medicine Society*, Vol. 20, No. 3, 2014, pp. 214–215. PMID: 2012755646

Schulz, V., "Single Dose of 900 mg of Hypericum Extract STW3-VI in Double Blind Study with 388 Moderately Severe Depressive Patients Is Equivalent to 20 mg Citalopram and Significantly Better Than Placebo [German]," *KIM: Komplementare und Integrative Medizin, Artztezeitschrift fur Naturheilverfahren*, Vol. 50, No. 4, 2009, pp. 71–72. PMID: CN-00754485

Willich, S. N., "[St. John's Wort Extract in Treatment of Moderately Severe Depression]," *Medizinische Monatsschrift für Pharmazeuten*, Vol. 30, No. 4, April 2007, pp. 151–154. PMID: 17484445

Reason Excluded: Excluded Comparator

Brattström, A., "Long-Term Effects of St. John's Wort (Hypericum Perforatum) Treatment: A 1-Year Safety Study in Mild to Moderate Depression," *Phytomedicine*, Vol. 16, No. 4, April 2009, pp. 277–283. PMID: 19299116

Sarris, J., "St. John's Wort and Kava in the Treatment of Major Depressive Disorder with Comorbid Anxiety," ClinicalTrials.gov, 2008. PMID: CN-00875859

Sarris, J., D. J. Kavanagh, G. Deed, and K. M. Bone, "St. John's Wort and Kava in Treating Major Depressive Disorder with Comorbid Anxiety: A Randomised Double-Blind Placebo-Controlled Pilot Trial," *Human Psychopharmacology*, Vol. 24, No. 1, January 2009, pp. 41–48. PMID: 19090505

Wan, H., and Y. Chen, "Effects of Antidepressive Treatment of Saint John's Wort Extract Related to Autonomic Nervous Function in Women with Irritable Bowel Syndrome," *International Journal of Psychiatry in Medicine*, Vol. 40, No. 1, 2010, pp. 45–56. PMID: 20565044

Reason Excluded: Excluded Participants

Harrer, G., U. Schmidt, and U. Kuhn, "Alternative Depressionbehandlung Mit Einem Hypericum-Extrakt," *Therapiewoche Neurologie Psychiatrie*, Vol. 5, 1991, pp. 710–716.

Hubner, W. D., S. Lande, and H. Podzuweit, "Hypericum Treatment of Mild Depressions with Somatic Symptoms," *Journal of Geriatric Psychiatry and Neurology*, Vol. 7, Suppl. 1, October 1994, pp. S12–S14. PMID: 7857500

Jackson, A., A. D'Avolio, G. Moyle, S. Bonora, G. Di Perri, L. Else, M. Simiele, G. J. Singh, D. Back, and M. Boffito, "Pharmacokinetics of the Co-Administration of Boceprevir and St John's Wort to Male and Female Healthy Volunteers," *Journal of Antimicrobial Chemotherapy*, Vol. 69, No. 7, July 2014, pp. 1911–1915. PMID: 24610312

Kasper, S., M. Gastpar, W. E. Muller, H. P. Volz, A. Dienel, M. Kieser, and H. J. Moller, "Efficacy of St. John's Wort Extract WS 5570 in Acute Treatment of Mild Depression: A Reanalysis of Data from Controlled Clinical Trials," *European Archives of Psychiatry and Clinical Neuroscience*, Vol. 258, No. 1, Feb, 2008, pp. 59–63. PMID: 18084790

König, C., "Hypericum Perforatum L. (Gemeines Johanniskraut) Als Therapeutikum Bei Depressiven Verstimmungszuständen - Eine Alternative Zu Synthetischen Arzneimitteln?" Dissertation Universität Basel, 1993.

Liu, Q., A. J. Lawrence, and J. H. Liang, "Traditional Chinese Medicine for Treatment of Alcoholism: From Ancient to Modern," *American Journal of Chinese Medicine*, Vol. 39, No. 1, 2011, pp. 1–13. PMID: WOS:000287004800001

Muller, J. E., I. Wentzel, L. Koen, D. J. H. Niehaus, S. Seedat, and D. J. Stein, "Escitalopram in the Treatment of Multisomatoform Disorder: A Double-Blind, Placebo-Controlled Trial," *International Clinical Psychopharmacology*, Vol. 23, No. 1, January 2008, pp. 43–48. PMID: WOS:000251907900006

Rapaport, M. H., A. A. Nierenberg, R. Howland, C. Dording, P. J. Schettler, and D. Mischoulon, "The Treatment of Minor Depression with St. John's Wort or Citalopram: Failure to Show Benefit Over Placebo," *Journal of Psychiatric Research*, Vol. 45, No. 7, July 2011, pp. 931–941.

Saito, Y. A., E. Rey, A. E. Almazar-Elder, W. S. Harmsen, A. R. Zinsmeister, G. R. Locke, and N. J. Talley, "A Randomized, Double-Blind, Placebo-Controlled Trial of St John's Wort for Treating Irritable Bowel Syndrome," *American Journal of Gastroenterology*, Vol. 105, No. 1, January 2010, pp. 170–177. PMID: 19809408

Reason Excluded: Excluded Intervention

Cipriani, A., C. Santilli, T. A. Furukawa, A. Signoretti, A. Nakagawa, H. McGuire, R. Churchill, and C. Barbui, "Escitalopram Versus Other Antidepressive Agents for Depression," *The Cochrane Database of Systematic Reviews*, No. 2, 2009, p. Cd006532. PMID: 19370639

Reason Excluded: Excluded Outcome

Whelan, A. M., T. M. Jurgens, and L. Lord, "Evaluating the Quality of Randomized Controlled Trials that Examine the Efficacy of Natural Health Products: A Systematic Review of

Critical Appraisal Instruments," *Evidence-Based Complementary and Alternative Medicine*, Vol. 6, No. 4, December 2009, pp. 441–448. PMID: WOS:000272179600004

Reason Excluded: Duplicate

Mannel, M., U. Kuhn, U. Schmidt, M. Ploch, and H. Murck, "St. John's Wort Extract LI160 for the Treatment of Depression with Atypical Features: A Double-Blind, Randomized, and Placebo-Controlled Trial," *Journal of Psychiatric Research*, Vol. 44, No. 12, September 2010, pp. 760–767. PMID: 20181361

Reason Excluded: Could Not Obtain

Bracher, A., "Johanniskraut 1x Täglich: Wirkung Klinisch Belegt/HYP611-Studie Bestätigt Wirksamkeit," *Verlagsbeilage Ärztliche Praxis*, Vol. 51, 2001, pp. 1–4.

Schmidt, U., "Zur Therapie Depressiver Verstimmungen," *Psychology,* Vol. 15, 1989, pp. 665–671.

Appendix D: Evidence Table of Included Studies

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Behnke et al., 2002 Country: NR Study Design: Multisite RCT, NR Purpose: a multicenter, randomized clinical comparison of a Hypericum extract and fluoxetine hydrochloride in the treatment of depression Quality Rating: Poor, one person missing from gender in Table 2, ITT mentioned but no results, unclear randomization, blinding, concealment	Number of Participants: 70 Diagnosis: Rating scale, MDD-ICD Comorbidities: NA Age (Years): 18–73 overall; 51.4 (SD 10.9) SJW; 48.0 (SD 12.6) fluoxetine Gender (% Male): 29% SJW; 34% fluoxetine (1 participant missing from SJW group) Inclusion Criteria: ability to provide written consent, age between 18 and 70, ICD-10 depression (category F32), and a score on the HAMD between 16 and 24 Exclusion Criteria: participation in a clinical study less than 4 weeks previously; pregnancy and lactation; insufficient contraception; suicide risk (HAMD score of 2, 3, or 4); dementia or other severe intellectual impairment; chronic alcohol abuse or dependence; chronic drug abuse or dependence; severe cardiac, liver, kidney, or respiratory insufficiency; neoplasia; Parkinson's or Alzheimer's disease; hypersensitivity to an ingredient of the Hypericum preparation; febrile illness; anemia; thyroid or parathyroid disease; and pituitary insufficiency	Extract: Hypericum perforatum Dosage: 150 mg (0.450–0.495 mg total hypericin) Co-interventions: NA Comparator: Fluoxetine Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 6 weeks Funding unclear, industry author, provided SJW	Depression Measures: • Remission (recovered rated as not ill or only borderline depressed on the Clinical Global Impression scale [CGI]), RR 0.94 (CI 0.65, 1.34) • Responder (≥50% decrease in total HAMD score), RR 0.84 (CI 0.560, 1.27) • Total HAMD, SMD −0.31 (CI −0.82, 0.19) Adverse Events: • Fluoxetine: Total mild to moderate adverse events 20 out of 35; Nausea/diarrhea 2 out of 35; Headache/dry mouth 2 out of 35 • SJW: Total mild to moderate adverse events 22 out of 35; Nausea/diarrhea 0 out of 34; Headache/dry mouth 0 out of 35

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Bernhardt, Liske, and Ebeling, 1993 Country: Germany Study Design: RCT Purpose: to collect further evidence for the application of hypericum Esbericum and to see if different dosages influence the effectiveness or onset of effect Quality Rating: Poor, no blinding of participants/ outcome assessments, information about follow-up insufficient, baseline outcomes unclear, statistical analysis not sufficiently reported	Number of Participants: 55 Diagnosis: Rating scale Comorbidities: NA Age (Years): 54.5 (SD 11.6) Gender (% Male): 29 Inclusion Criteria: Mild to moderate depression (as measured by the HAMD) Exclusion Criteria: NA	Extract: Hypericin Dosage: 0.25 mg extract, 3 times per day (morning/noon/night) for 4 weeks; 0.25 mg extract, 3 times per day (2 in the morning, 1 at noon), for 4 weeks; Co-interventions: NA Primary Endpoint: NA Power Calculation: No Funding unclear, not reported	Depression Measures: • Results show a preference for trice daily dosage

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Bjerkenstedt et al., 2005 Country: Sweden Study Design: Multisite RCT, NR Purpose: to investigate the efficacy and tolerability of LI 160, a SJW dry extract, in mild to moderate depression as compared with fluoxetine and placebo in a prospective, randomized, double-blind, double-dummy parallel cohort trial Quality Rating: Fair, unclear randomization and allocation, adequate blinding	Number of Participants: 174 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): 49.1 (SD 12.0) SJW; 50.4 (SD 11.6) fluoxetine; 51.4 (SD 11.8) placebo Gender (% Male): 20% (SJW); 24% (fluoxetine); 18% (placebo) Inclusion Criteria: Caucasian females and males, age 18 to 70 years; a minimum total score of 21 on the 21-item HAMD; history of at least two episodes of non-psychotic MDD; capacity and willingness to give informed consent and to comply with study procedures. Exclusion Criteria: a diagnosis of psychotic mental disorder; other disorders requiring concomitant psychoactive medication; monoamine oxidase (MAO) inhibitor treatment within 14 days prior to entry; history of treatment-resistant MDD (at least two different antidepressants over 6 weeks at sufficient doses) from at least two previous depressive episodes; risk of suicide; history of seizure disorder; alcohol or substance abuse; other serious unstable acute or chronic medical illness; severely impaired hepatic or renal function; pregnancy, breast-feeding, or use of inadequate contraceptives in fertile women; known intolerance or hypersensitivity to study medications; substantial placebo response (HAMD reduction ≥20%) at the end of placebo run-in phase; treatment with any investigational drug during three months prior to inclusion; participation in another clinical trial within 30 days before start of the study.	Extract: Hypericum LI 160 Dosage: 300 mg, 3 times per day, daily, 6 weeks Co-interventions: NA Comparator: Fluoxetine, placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 4 weeks No industry funding	Depression Measures: • Remission (HAMD total score <8), RR 0.87 (CI 0.46, 1.64) • Remission (HAMD total score <8), RR 3.31 (CI 1.149, 9.52) • Response (≥50% decrease in total score since baseline), RR 1.07 (CI 0.67, 1.7) • Response (≥50% decrease in total score since baseline), RR 1.100 (CI 0.68, 1.77) • total HAMD, SMD −7.00 (CI −0.44, 0.31) • total HAMD, SMD 0.01 (−0.37, 0.39) Adverse Events: • Fluoxetine: Total AEs 52 out of 56; Skin and appendages disorders 5 out of 56; Psychiatric disorders 8 out of 56; Metabolic and nutritional disorders 6 out of 56; GI system disorders 17 out of 56; Central and peripheral nervous system disorders 3 out of 56; Body as a whole 18 out of 56; Autonomic nervous system disorders 12 out of 56; Adverse reaction “definitely, probably, or possibly” related to study drug 39 out of 56 • Placebo: Total AEs 27 out of 57; Skin and appendages disorders 3 out of 57; Psychiatric disorders 3 out of 57; Metabolic and nutritional disorders . out of 57; GI system disorders 11 out of 57; Central and peripheral nervous system disorders 4 out of 57; Body as a whole 5 out of 57; Autonomic nervous system disorders 8 out of 57; Adverse reaction “definitely, probably, or possibly” related to study drug 15 out of 57 • SJW: Total AEs 38 out of 57; Skin and appendages disorders 9 out of 57; Psychiatric disorders 2 out of 57; Metabolic and nutritional disorders . out of 57; GI system disorders 6 out of 57; Central and peripheral nervous system disorders 10 out of 57; Body as a whole 13 out of 57; Autonomic nervous system disorders 10 out of 57; Adverse reaction “definitely, probably, or possibly” related to study drug 24 out of 57

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Brenner et al., 2000 Country: United States Study Design: RCT Purpose: to compare an extract of hypericum with the SSRI sertraline in a controlled trial to establish preliminary efficacy and tolerability data in depressed patients Quality Rating: Fair, unclear randomization, allocation concealment, blinding	Number of Participants: 30 Diagnosis: MDD-DSM, Rating scale, Other diagnosis, must have both Comorbidities: NA Age (Years): 45 Gender (% Male): 37% Inclusion Criteria: male or female outpatients aged 18 to 65 years with a score of >17 on the HAMD and a DSM-IV diagnosis of MDD (single or recurrent episodes), dysthymic disorder, adjustment disorder with depressed mood, or depressive disorder not otherwise specified. Exclusion Criteria: Pregnant women or women not using medically accepted means of birth control were excluded, as were patients with severe depression and a history of attempted suicide or acute suicidal state, schizophrenia or marked agitation requiring additional medication, or chronic alcohol or drug dependency. Also excluded were patients who had failed to respond to adequate trials of an antidepressant drug, who had received an investigational drug within 4 weeks before beginning the study, or who had been treated with hypericum or sertraline previously. Patients with mental retardation or emotional or intellectual difficulties that could invalidate informed consent or limit their ability to comply with the study protocol were also excluded. Patients whose HAMD scores improved by >20% between screening and baseline were excluded from randomization.	Extract: LI 160 Dosage: 600 mg per day during week 1, followed by 900 mg per day for the remainder of the trial Co-interventions: NA Comparator: Sertraline Primary Endpoint: HAMD, CGI global severity Power Calculation: Insufficient power (posthoc analysis) Follow-Up Time: 7 weeks No industry funding	Depression Measures: • HAMD, SMD 0.03 (CI −0.71, 0.77) • Clinical response (≥50% reduction HAMD), RR 1.35 (CI 0.61, 2.99) Adverse Events: • SJW total AEs: 3 events in 2 participants out of 15 • Sertraline total AEs: 4 events in 2 participants out of 15

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Fava et al., 2005 Country: United States Study Design: Multisite RCT, 2 Purpose: to compare the antidepressant efficacy and safety of a standardized extract of SJW with both placebo and fluoxetine Quality Rating: Poor, severity not described, dropouts of about 50%	Number of Participants: 135 Diagnosis: MDD-DSM, Rating scale, Other diagnosis, Severe Combined Immunodeficiency Comorbidities: NA Age (Years): 37.3 (SD 11.0) Gender (% Male): 43% Inclusion Criteria: either sex, any ethnic origin, Age 18 to 65 years, Current experience of a major depressive episode according to DSM-IV of at least 2 weeks' duration, HAMD-17 total score of ≥16 at both screen and baseline, Negative pregnancy test within 5 days before study start in women of childbearing potential (nonchildbearing potential was defined as postmenopause for at least 1 year or surgical sterilization or hysterectomy at least 3 months before study start), Use of adequate contraception in women of childbearing potential, Readiness and ability on the part of the patient to comply with the physician's instructions and to fill out the self-report measures in connection with their examination at the study visits, Written informed consent. Exclusion Criteria: Pregnancy, lactation, or nonuse of medically accepted means of contraception in women of childbearing potential; Current serious suicidal or homicidal risk (according to investigator's judgment); Serious or unstable medical illness including cardiovascular, hepatic, renal, respiratory, endocrine, neurological, or hematologic disease; History of seizure disorder; One or more of the following DSM-IV diagnoses: organic mental disorders, substance use disorders (including alcohol) active within the last 6 months, schizophrenia, delusional disorder, psychotic disorders not elsewhere classified, bipolar disorder, or antisocial personality disorder; History of multiple adverse drug reactions or allergy to the study drugs; Mood-congruent or mood-incongruent psychotic features; Any of the following treatments at baseline or within the specified time frame before baseline: other psychotropic drugs (14 days), other investigational psychotropic drug (40 days) fluoxetine (40 days), or any	Extract: LI-160 Dosage: 300 mg, 3 times a day Co-interventions: NA Comparator: Fluoxetine, placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 12 weeks No industry funding	Depression Measures: • HAMD, SMD −0.37 (CI −0.79, 0.06) • HAMD, SMD −0.44 (CI −0.85, −0.03) • Remission (HAMD <8), RR 1.27 (CI 0.71, 2.259) • Remission (HAMD <8), RR 1.8 (CI 0.9, 3.6) Adverse Events: • Fluoxetine: URTI 5 out of 47; Sleepiness 6 out of 47; Rash 0 out of 47; Nausea 4 out of 47; Muscle Pain/Aches 4 out of 47; Joint Pain 2 out of 47; Insomnia 5 out of 47; Headache 12 out of 47; Gastrointestinal Tract Upset 10 out of 47; Flu 1 out of 47; Dry Mouth 6 out of 47; Diarrhea 7 out of 47; Cold Symptoms 7 out of 47 • Placebo: URTI 6 out of 43; Sleepiness 3 out of 43; Rash 5 out of 43; Nausea 7 out of 43; Muscle Pain/Aches 3 out of 43; Joint Pain 7 out of 43; Insomnia 6 out of 43; Headache 12 out of 43; Gastrointestinal Tract Upset 5 out of 43; Flu 1 out of 43; Dry Mouth 4 out of 43; Diarrhea 4 out of 43; Cold Symptoms 4 out of 43 • SJW: URTI 5 out of 45; Sleepiness 8 out of 45; Rash 0 out of 45; Nausea 9 out of 45; Muscle Pain/Aches 5 out of 45; Joint Pain 4 out of 45; Insomnia 7 out of 45; Headache 19 out of 45; Gastrointestinal Tract Upset 9 out of 45; Flu 5 out of 45; Dry Mouth 10 out of 45; Diarrhea 3 out of 45; Cold Symptoms 6 out of 45

Study Details	Participants	Intervention/Treatment	Outcomes/Results
	other investigational drug (1 month); Unacceptability to discontinue or likelihood to need medication that is prohibited as concomitant treatment during the study; Clinical or laboratory evidence of hypothyroidism; Failure to respond during the course of current major depressive episode to at least 2 adequate antidepressant trials, defined as 8 weeks or more of treatment with either imipramine 150 mg or greater (or its tricyclic equivalent), phenelzine 60 mg or greater (or its MAO inhibitor equivalent), or fluoxetine 20 mg or greater (or its selective serotonin reuptake inhibitor equivalent); Any other condition that, in the investigator's judgment, may pose a significant risk to the patient's health or may decrease the chances of obtaining reliable data to achieve the objectives of the study; Mental condition rendering the patient unable to understand the nature, scope, and possible risks of the study; History or suspicion of unreliability, poor cooperation, or noncompliance with medical treatment.		

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Gastpar, Singer, and Zeller, 2005 Country: Germany Study Design: Multisite RCT, 21 Purpose: to demonstrate the non-inferiority of hypericum extract versus sertraline in the treatment of moderate depression Quality Rating: Poor, not really ITT analysis	Number of Participants: 241 Diagnosis: MDD-DSM, Rating scale, MDD-ICD Comorbidities: NA Age (Years): 48.3 (SD 12.7) SJW; 49.5 (SD 13.8) Sertraline Gender (% Male): 20.6 SJW; 30.6 Sertraline Inclusion Criteria: written consent after comprehensive explanation of the content, significance, and scope of the clinical trial by the investigator; age of 18–70 years; females taking adequate contraceptive or without childbearing potential; depression with a score of 20–24 on the HAMD (items1–17); and diagnosis of moderate depression (depressive episode or recurrent depressive disorder) defined by ICD-10 F32.1 or F33.1, respectively, with four or more typical symptoms for depressive episodes according to DSM-IV major depressive episode (296.2x) and recurrent major depression (296.3x). Exclusion Criteria: resistance to treatment; schizophrenia, psychosis, or dementia; depression due to a serious general medical cause; known hypersensitivity to hypericum extract or sertraline; known photosensitivity; specific antidepressant psychotherapy during the last two months or treatment, with antidepressants during the last six weeks; and suicidal tendency determined by scores of > 2 in item 3 of HAMD or known attempted suicide.	Extract: STW3 Dosage: 612 mg per day, for 12 weeks Co-interventions: NA Comparator: Sertraline Primary Endpoint: HAMD Power Calculation: Yes Follow-Up Time: 24 weeks Funding unclear, industry author, provided SJW	Depression Measures: • HAMD, SMD −0.25 (CI −0.5, 0.0) • Relapse (significant increase in HAMD score of more than 100% (with a score of at least 10 as last study value) or a score of at least 16 points in follow-up phase), RR 0.24 (CI 0.03, 2.11) • Responder (HAMD<10 or decrease ≥50% in total HAMD score), RR 0.93 (CI 0.79, 1.090) Adverse Events: • SJW: Total AEs 189 out of 102; Skin and integumentary system 2 out of 102; Psychiatric disorders 0 out of 102; Generalized disturbances (fatigue) 1 out of 102; Diseases of live and hepatic duct 1 out of 102; Digestive Tract 9 out of 102; Central and peripheral nervous system 1 out of 102 • Sertraline: Total AEs 112 out of 98; Skin and integumentary system 0 out of 98; Psychiatric disorders 5 out of 98; Generalized disturbances (fatigue) 2 out of 98; Diseases of liver and hepatic duct 0 out of 98; Digestive Tract 7 out of 98; Central and peripheral nervous system 4 out of 98

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Gastpar, Singer, and Zeller, 2006 Country: Germany Study Design: Multisite RCT, 21 Purpose: to demonstrate for the first time the comparable efficacy of the hypericum extract STW3-VI to citalopram after short-term treatment of patients with moderate depression with a once-daily dose regimen Quality Rating: Good, adequate randomization and double dummy blinding, performed ITT analysis, 80% follow-up	Number of Participants: 388 Diagnosis: MDD-DSM, Rating scale, MDD-ICD Comorbidities: NA Age (Years): SJW 50.8 (SD 12.1); Citalopram 49.3 (SD 10.7); Placebo 49.4 (SD 12.7) Gender (% Male): SJW 34.4, Citalopram 35.4, Placebo 26.9 Inclusion Criteria: written consent after comprehensive explanation of the content, significance and scope of the clinical trial by the investigator; age: 18–70 years; females taking adequate contraceptive or without childbearing potential; patients having depression with a score of 20–24 on the HAMD (items 1–17); and diagnosis of moderate depression (first manifestation or recurrent depressive disorder) defined by ICD-10 F32.1 or F33.1 according to DSM-IV major depressive episode (296.2x) and recurrent major depression (296.3x). Exclusion Criteria: diagnosis of resistance to depression treatment; known schizophrenia; psychosis or dementia; depressive mood due to a serious general disease; known hypersensitivity to study medication; known photosensitivity; specific antidepressant psychotherapy during the last two months or treatment with psychoactive drugs (antidepressants, neuroleptic drugs, antidementive drugs, anxiolytic drugs, etc.) during the last 3 weeks (6 weeks for fluoxetine) prior to study enrollment; and determined suicidal tendency by scores of >2 in item 3 of HAMD or known attempted suicide.	Extract: STW3-VI Dosage: 900 mg per day, for 6 weeks Co-interventions: NA Comparator: Citalopram, placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 6 weeks Funding unclear, industry author, provided SJW	Depression Measures: • HAMD, SMD −0.4 (CI −0.65, −0.16) • HAMD, SMD 0 (CI −0.24, 0.24) • Responder (HAMD< 10 or decrease >=50% in total HAMD score), RR 0.97 (CI 0.78, 1.21) • Responder (HAMD< 10 or decrease >=50% in total HAMD score), RR 1.38 (CI 1.06, 1.8) Adverse Events: • Citalopram: Withdrawal 11 out of 127; Total AEs 94 out of 127; Skin and subcutaneous tissue disorders 6 out of 127; Severe intensity of AEs 3 out of 127; Serious AEs 2 out of 127; Nervous system disorders 9 out of 127; Musculoskeletal and connective tissue disorders 5 out of 127; Infections and infestations 17 out of 127; Gastrointestinal disorders 23 out of 127; Ear and labyrinth disorders 11 out of 127 • Placebo: Withdrawal 6 out of 130; Total AEs 70 out of 130; Skin and subcutaneous tissue disorders 3 out of 130; Severe intensity of AEs 1 out of 130; Serious AEs 4 out of 130; Nervous system disorders 10 out of 130; Musculoskeletal and connective tissue disorders 3 out of 130; Infections and Infestations 17 out of 130; Gastrointestinal disorders 20 out of 130; Ear and labyrinth disorders 6 out of 130 • SJW: Withdrawal 4 out of 133; Total AEs 58 out of 133; Skin and subcutaneous tissue disorders 4 out of 133; Severe intensity of AEs 1 out of 133; Serious AEs 0 out of 133; Nervous system disorders 1 out of 133; Musculoskeletal and connective tissue disorders 6 out of 133; Infections and infestations 20 out of 133; Gastrointestinal disorders 11 out of 133; Ear and labyrinth disorders 2 out of 133

Study Details	Participants	Intervention/Treatment	Outcomes/Results
HDTSG, 2002 Country: United States Study Design: Multisite RCT, 12 Purpose: to test the efficacy and safety of a well-characterized hypericum extract (LI-160) in major depressive disorder Quality Rating: Fair, <80% completed but did primary outcome analysis	Number of Participants: 338 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): SJW 43.1 (SD 13.5); Sertraline 43.9 (SD 13.9); Placebo 40.1 (SD 12.2) Gender (% Male): SJW 35.4; Sertraline 33.3; Placebo 33.6 Inclusion Criteria: age at least 18 years; current diagnosis of major depression; minimum total score of 20 on the 17-item HAMD and a maximum score of 60 on the Global Assessment of Functioning at screening and baseline following a 1-week, single-blind, placebo run-in; no more than a 25% decrease in HAMD total score between screening and baseline; capacity to give informed consent and follow study procedures; and identification of a close personal contact to be notified if warranted by clinical concerns. Exclusion Criteria: a score above 2 on the HAMD suicide item; attempted suicide in the past year or current suicide or homicide risk; being pregnant, planning pregnancy, breastfeeding, or not using medically acceptable birth control; clinically significant liver disease or liver enzyme levels elevated to at least twice the upper normal limit; serious mental illness; history of seizure disorder; Severe Combined Immunodeficiency diagnoses indicating alcohol or other substance abuse disorder in the past 6 months or lifetime diagnoses of schizophrenia, schizoaffective or other psychotic disorder, bipolar disorder, panic disorder, or obsessive-compulsive disorder; history of psychotic features of affective disorder; evidence of untreated or unstable thyroid disorder; no response to at least 2 adequate trials of antidepressants in any depressive episode; daily dose of hypercium or sertraline for at least 4 weeks within the past 6 months; current use of other psychotropic drugs, other medicines, dietary supplements, natural remedies, or botanical preparations with psychoactive properties; use of investigational drugs within 30 days of baseline or of other psychotropic drugs within 21 days of baseline (within 6 weeks for fluoxetine); allergy	Extract: LI-160 Dosage: 300 mg, 3 times a day, for 8 weeks Co-interventions: NA Comparator: Sertraline, placebo Primary Endpoint: HAMD Power Calculation: Insufficient power (posthoc analysis) Follow-up Time: 8 weeks No industry funding	Depression Measures: • HAMD total score, SMD 0.08 (CI −0.17, 0.34) • HAMD total score, SMD 0.3 (CI 0.04, 0.569) • Responder (any = full + partial above), RR 0.78 (CI 0.579, 1.06) • Responder (any = full + partial above), RR 0.88 (CI 0.64, 1.21) Adverse Events: • Placebo: Withdrawals due to AEs 3 out of 116; Swelling 9 out of 116; Sweating 14 out of 116; Nausea 24 out of 116; Frequent urination 13 out of 116; Forgetfulness 26 out of 116; Diarrhea 22 out of 116; Anorgasmia 16 out of 116 • SJW: Withdrawals due to AEs 2 out of 113; Swelling 21 out of 112; Sweating 20 out of 112; Nausea 21 out of 112; Frequent urination 30 out of 112; Forgetfulness 28 out of 112; Diarrhea 23 out of 112; Anorgasmia 28 out of 112 • Sertraline: Withdrawals due to AEs 5 out of 111; Swelling 9 out of 111; Sweating 32 out of 111; Nausea 41 out of 111; Frequent urination 23 out of 111; Forgetfulness 13 out of 111; Diarrhea 42 out of 111; Anorgasmia 35 out of 111

Study Details	Participants	Intervention/Treatment	Outcomes/Results
	or hypersensitivity to study medications; positive urine drug screen; introduction of psychotherapy within 2 months of enrollment or any psychotherapy specifically designed to treat depression; and mental retardation or cognitive impairment.		

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Hänsgen, Vesper, and Ploch, 1994 Country: Germany Study Design: Multisite RCT, 11 Purpose: patients were treated in a double-blind study for a period of 6 weeks either with hypericum extract LI 160 or with placebo Quality Rating: Poor, no ITT analysis	Number of Participants: 108 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): 53.0 (SD 7.5) SJW; 53.5 (SD 10.3) Placebo Gender (% Male): 42 SJW; 32 Placebo Inclusion Criteria: Male and female subjects, aged 18 to 70 years, fulfilled the criteria for major depression according to DSM-III-R. Further inclusion criteria were a total score on the HAMD of 16 or more and a duration of their depressive episode between a minimum of 2 weeks and a maximum of 6 months. Exclusion Criteria: psychotic features, suicide risks, severe medical illnesses, dependent on alcohol, drugs, or medications, pregnant or inadequate contraception, or being treated concomitantly with other psychotropic drugs.	Extract: LI 160 Dosage: 300 mg, 3 times a day, for 6 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD; von Zerssen test Depression Scale(D-S); Hansgens complaint inventory (BEB); CGI Power Calculation: No Follow-Up Time: 5 and 6 weeks Funding unclear, industry author, provided SJW	Depression Measures: • Responder HAMD, RR 3.03 (CI 1.77, 5.17) Adverse Events: • Placebo: mild AEs 2 out of 34; gastrointestinal symptoms 2 out of 34; Sleep disturbance 0 out of 34 • SJW: mild AEs 1 out of 33; gastrointestinal symptoms 0 out of 33; Sleep disturbance 1 out of 33

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Harrer, Hübner, and Podzuweit, 1994 Country: Germany Study Design: Multisite RCT, 6 Purpose: undertake a comparison between the effectiveness and tolerance of the standardized hypericum extract LI 160 and maprotiline in six specialist centers Quality Rating: Poor, mostly unclear methods, adequate blinding, no ITT analysis	Number of Participants: 102 Diagnosis: Rating scale, MDD-ICD Comorbidities: NA Age (Years): SJW 43.8 (SD 13.4); Maprotiline 47.6 (SD 10.9) Gender (% Male): SJW 25; Maprotiline 31 Inclusion Criteria: Male and female patients, aged 24 to 65 years; depression according to ICD-10, F 32.1 (single, moderately severe depressive episode with usually dejected mood, loss of interest, loss of happiness, and reduction in drive) for at least 2 weeks. The raw sum score in the HAMD in the version with 17 items had to be at least 16. Exclusion Criteria: Previous treatment with psychopharmacologic agents must have ended at least 4 weeks prior to the study. An allergy to sunlight and, in relation to the known adverse reactions of maprotiline, conduction disorders in the heart, narrow-angle glaucoma, and adenoma of the prostate were also exclusion criteria, in addition to those normally applied in clinical trials.	Extract: LI 160 Dosage: 300mg, 3 times a day, for 4 weeks Co-interventions: NA Comparator: Maprotiline Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 4 weeks Funding unclear, NR	Depression Measures: • Remission (no longer ill), RR 1.83 (CI 1.02, 3.3) • Responder (HAMD <10 or decrease ≥50% in total HAMD score), RR 0.96 (CI 0.67, 1.38) Adverse Events: • LI 160: Total AEs 25 out of 44; Tiredness 2 out of 44; Other 4 out of 44; Heart complaints 0 out of 44; Gastrointestinal symptoms 8 out of 44; Exanthema 0 out of 44; Dryness of mouth 3 out of 44; Dizziness, confusion 8 out of 44 • Maprotiline: Total AEs 44 out of 42; Tiredness 11 out of 42; Other 5 out of 42; Heart complaints 3 out of 42; Gastrointestinal symptoms 8 out of 42; Exanthema 2 out of 42; Dryness of mouth 7 out of 42; Dizziness, confusion 8 out of 42

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Harrer et al., 1999 Country: NR Study Design: Multisite RCT, 17 Purpose: the antidepressant efficacy of a daily dose of 800 mg of the SJW extract LoHyp-57 (dry extract of SJW, drug extract ratio 5-7:1, solvent, ethanol 60% [weight/ weight]) was shown to be equivalent to that of 20 mg fluoxetine (CAS 54910-89-3) in elderly patients with mild or moderate depressive episodes according to ICD 10. Quality Rating: Fair, ITT analysis, randomized and double blind but unclear how	Number of Participants: 228 Diagnosis: MDD-ICD Comorbidities: NA Age (Years): SJW 68.4; Fluoxetine 69.1 Gender (% Male): 13.4 Inclusion Criteria: Male and female patients aged 60 to 80 years, suffering their first psychiatric illness, with symptoms satisfying the diagnostic guidelines of F 32.0 and F 32.1 according to ICD 10, were to be included in the study. Exclusion Criteria: A demential disorder with a score of ≤25 on the Mini Mental Status Test.	Extract: LoHyp-57 Dosage: 400 mg, 2 times a day, for 6 weeks Co-interventions: NA Comparator: Fluoxetine Primary Endpoint: HAMD Power Calculation: Yes Follow-Up Time: 6 weeks Funding unclear, NR	Depression Measures: HAMD • Remission (completely resolved), RR 1.139 (CI 0.42, 3.07) • Responder (global HAMD ≤10 points or decrease in 50% by visit 1), RR 0.99 (CI 0.81, 1.21) Adverse Events: • Total LoHyp: 30 AEs in 12 participants out of 69; 1 with intermittent abdominal pain/stomachache; 1 with recurrent biliary colic; 1 with racing heartbeat at night; 1 with tachycardia lasting 2 hours; 1 with nausea, diarrhea, swelling of face and pruritis; 1 with vertigo, nausea, misty vision; 1 with sensation of rising heat and sweating; 1 with dry mouth and disturbed sleep; 1 with stabbing chest pains, headaches, vertigo; 1 with excessive rise in blood pressure, sleep difficulty, chest pressure, general malaise; 1 with nausea, inner restlessness, vomiting; 1 with anxiety attacks, loss of appetite, sleep disturbance; 1 with fatigue • Total fluoxetine: 42 AEs in 17 participants out of 68; 1 with diarrhea, nausea, abdominal symptoms; 1 with restlessness, palpitations; 1 with constipation, muscle and skin tension, disturbances of concentration and thought; 1 with nausea, mild vertigo; 1 with night sweats; 1 with restlessness; 1 with stomatitis, increased restlessness;1 with racing heartbeat, rise in blood pressure; 1 with fatigue; 1 with sleep disturbance, tremor, restlessness, headache, loss of appetite; 1 with morning fatigue, little drive; 1 with feeling of fullness, dry mouth, constipation, allergic eczema; 1 with vertigo; 1 with nausea, restlessness; 1 with nausea, headache, diarrhea, increased restlessness, stomachache; 1 with upper abdominal symptoms; 1 with nausea headache, tachycardia

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Kalb, Trautmann-Sponsel, and Kieser, 2001 Country: Germany Study Design: Multisite RCT, NR Purpose: to demonstrate the efficacy of hypericum extract WS 5572 versus placebo and investigate its tolerability and safety in patients suffering from mild or moderate depressive disorders according to DSM-IV Quality Rating: Good, achieved adequate randomization and double-blinding, all participants completed	Number of Participants: 72 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): SJW 48 (SD 11); Placebo 49 (SD 10) Gender (% Male): SJW 29.7; Placebo 37.1 Inclusion Criteria: male and female outpatients aged between 18 and 65 years; total score for the HAMD (17-item version) >16 at study entry and during a subsequent baseline investigation (3 to 7 days later); diagnosis of mild or moderate MDD with single or recurrent episodes according to DSM-IV criteria (diagnostic codes 296.21, 296.31, 296.22, or 296.32). The diagnosis of depression had to be confirmed by a psychiatrist when a patient was recruited by non-psychiatrists. Exclusion Criteria: suicidal tendency (known attempted suicide or item 3 of the HAMD (suicide) >2 points); organic brain syndrome; major psychiatric diseases (other than depression); disorders caused by psychotropic substances; pre-treatment with fluoxetine during the last 6 weeks, with paroxetine or doxepin during the last 2 weeks before baseline; concomitant medication with other antidepressants, psychotropic drugs, or reserpine; severe metabolic, internal, or neoplastic diseases; substance abuse, pregnancy, or lactation period. Concomitant medication doses required for the treatment of non-psychiatric conditions were to be maintained unchanged during the course of the study where possible.	Extract: WS 5572 Dosage: 300 mg/3 times a day/6 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 6 weeks Industry funding	Depression Measures: • HAMD total score, SMD −0.96 (CI −1.45, −0.47) • Responder (reduction of ≥50%), RR 1.45 (CI 0.92, 2.29) Adverse Events: • Placebo: Total AEs 2 out of 35; Sinusitis 0 out of 35; Gastroenteritsis 1 out of 35; Common cold 0 out of 35; Bronchitis 1 out of 35 • SJW: Total AEs 3 out of 34; Sinusitis 1 out of 34; Gastroenteritsis 0 out of 34; Common cold 1 out of 34; Bronchitis 1 out of 34

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Kasper et al., 2006 Country: Germany Study Design: Multisite RCT, 16 Purpose: to assess the antidepressant efficacy and safety of SJW extract WS 5570 at doses of 600 mg/day (given only once daily) and 1200 mg/day (given as 600 mg twice daily) over 6 weeks of treatment in patients suffering from a major depressive episode Quality Rating: Fair, adequate randomization and double-blinding, ITT analysis	Number of Participants: 332 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): SJW 600 mg 46.3 (SD 11.5); SJW 1200 mg 46.1 (SD 10.7); Placebo 46.9 (SD 11.8) Gender (% Male): SJW 600 mg 43.7; SJW 1200 mg 33.9; Placebo 30.9 Inclusion Criteria: The episode was required to be of at least two weeks' but not more than one year's duration, male and female patients, 18 to 65 years of age, with a diagnosis of a mild or moderate, single or recurrent, major depressive episode as defined by the DSM-IV (296.21, 296.31; 296.21 or 296.22, 296.31 and 296.32). Participants were required to have HAMD total score ≥18 and HAMD item "depressive mood" ≥2 at baseline. Exclusion Criteria: NA	Extract: WS 5570 Dosage: 600 or 1200 mg/day, for 6 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD Power Calculation: Yes Follow-Up Time: 6 weeks Funding unclear, industry author, provided SJW	Depression Measures: • Remission (HAMD ≤7), RR 2.72 (CI 1.55, 4.79) • Responder (≥50% decrease in total HAMD score), RR 2.17 (CI 1.55, 3.05) • Total HAMD, SMD −0.82 (CI −1.11, −0.53) • Quality of life SF-36 mental health, SMD 0.5 (CI 0.22, 0.78) • Quality of life SF-36 physical health, SMD 0.23 (CI −0.05, 0.51) Adverse Events: • Placebo: Vascular disorders 0 out of 82; Skin and subcutaneous disorders 4 out of 82; Serious AEs 0 out of 82; Respiratory, thoracic, and medistinal disorders 2 out of 82; Reproductive system and breast disorders 0 out of 82; Renal and urinary disorders 0 out of 82; Psychiatric disorder 0 out of 82; Nervous system disorder 2 out of 82; Musculoskeletal and connective tissue disorders 1 out of 82; Metabolism and nutrition disorders 1 out of 82; Investigations 0 out of 82; Injury, poisoning, and procedural complications 1 out of 82; Infections and infestations 2 out of 82; General disorders and administration site conditions 0 out of 82; Gastrointestinal disorders 13 out of 82; Eye disorders 0 out of 82; Ear and labyrinth disorders 1 out of 82; All AEs 22 out of 82; AE Potentially related to the treatment 13 out of 82 • WS 5570: 1200 mg Vascular disorders 1 out of 127; Skin and subcutaneous disorders 2 out of 127; Serious AEs 2 out of 127; Respiratory, thoracic, and medistinal disorders 5 out of 127; Reproductive system and breast disorders 2 out of 127; Renal and urinary disorders 0 out of 127; Psychiatric disorder 2 out of 127; Nervous system disorder 6 out of 127; Musculoskeletal and connective tissue disorders 2 out of 127; Metabolism and nutrition disorders 1 out of 127; Investigations 0 out of 127; Injury, poisoning, and procedural complications 1 out of 127; Infections and infestations 4 out of 127; General disorders and administration site conditions 2 out of 127;

Study Details	Participants	Intervention/Treatment	Outcomes/Results
			Gastrointestinal disorders 30 out of 127; Eye disorders 1 out of 127; Ear and labyrinth disorders 2 out of 127; All AEs 50 out of 127; AE Potentially related to the treatment 31 out of 127 • WS 5570: 600 mg Vascular disorders 1 out of 123; Skin and subcutaneous disorders 4 out of 123; Serious AEs 1 out of 123; Respiratory, thoracic, and medistinal disorders 4 out of 123; Reproductive system and breast disorders 1 out of 123; Renal and urinary disorders 1 out of 123; Psychiatric disorder 2 out of 123; Nervous system disorder 6 out of 123; Musculoskeletal and connective tissue disorders 1 out of 123; Metabolism and nutrition disorders 1 out of 123; Investigations 1 out of 123; Injury, poisoning and procedural complications 1 out of 123; Infections and infestations 7 out of 123; General disorders and administration site conditions 2 out of 123; Gastrointestinal disorders 24 out of 123; Eye disorders 0 out of 123; Ear and labyrinth disorders 3 out of 123; All AEs 49 out of 123; AE Potentially related to the treatment 30 out of 123

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Kasper et al., 2008 Country: Germany and Sweden Study Design: Crossover RCT Purpose: the efficacy and safety of hypericum extract WS 5570 in preventing relapse during 6 months' continuation treatment and 12 months' long-term maintenance treatment after recovery from an episode of recurrent depression Quality Rating: Fair, ITT analysis but less than 80% follow-up during continuation phase	Number of Participants: 570 Diagnosis: MDD-DSM, Rating scale, MDD-ICD Comorbidities: NA Age (Years): 47.5 (SD 10.7); Placebo 47.4 (SD 11.8) Gender (% Male): SJW 27; Placebo 24.3 Inclusion Criteria: 18–65 years old, had to suffer from a recurrent episode of major depression (ICD-10 F33.0 or F33.1, and DSM-IV 296.3), and had to have a history of 2 or 3 previous episodes according to ICD-10 and DSM-IV criteria. HAMD 17 total score ≥20 points and ≥2 points for item "depressive mood" were required at screening and at the start of acute treatment. Exclusion Criteria: Anyone with a diagnosis of schizophrenia, acute anxiety disorder, adjustment disorder, chronic or psychotic depression, bipolar disorder, acute posttraumatic stress disorder, or substance abuse (except nicotine and caffeine) were excluded. Patients with increased risk of suicide (e.g., HAMD item "suicide" ≥2) or previous attempted suicide were not allowed to participate, and concomitant medical and nonmedical antidepressant treatments were prohibited.	Extract: WS 5570 Dosage: 300 mg, 3 times a day, for 26 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: Relapse during continuation treatment was defined by any of the following observations: a HAMD total score ≥16 points, or clinical diagnosis of a depressive episode according to ICD-10 criteria, or premature treatment termination related to lack of efficacy (as determined by the physician or the patient) Power Calculation: Yes Follow-Up Time: 32 weeks Industry funding	Depression Measures: • HAMD, SMD −0.21 (CI −0.42, −0.01) • Relapse (HAMD ≥16 or clinical diagnosis of depressive episode ICD-10 or premature treatment termination related to lack of efficacy), RR 0.7 (CI 0.48, 1.02) Adverse Events: • Acute phase WS 5570: Withdrawal due to AEs 13 out of 703; Total AEs 221 out of 703 • Continue Phase Placebo: Withdrawal due to AEs 6 out of 194; Total AEs 213 out of 194 • Continue Phase WS 5570: Withdrawal due to AEs 8 out of 376; Total AEs 317 out of 376 • Prophylaxis Phase Placebo: Withdrawal due to AEs 2 out of 136 • Prophylaxis Phase WS 5570: Withdrawal due to AEs 1 out of 138

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Laakmann, Dienel, and Kieser, 1998 Country: Germany Study Design: Multisite RCT, 11 Purpose: investigate clinical significance of the hyperforin content for the efficacy of hypericum extracts, depending on the severity of the patients' depression Quality Rating: Good	Number of Participants: 196 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): SJW WS 5572 47.3 (SD 11.8); SJW WS 5573 48.7 (SD 11.8); Placebo 51.0 (SD 12.7) Gender (% Male): SJW WS 5572 18.4; SJW WS 5573 14.3; Placebo 28.6 Inclusion Criteria: male and female outpatients suffering from mild or moderate depression according to DSM-IV criteria (either single or recurrent episode), between 18 and 65 years of age, and an initial score ≥17 on the HAMD, (17-item version). Exclusion Criteria: risk of suicide or a score of ≥2 on HAMD item 3 (suicidality); organic brain syndrome; compulsive, schizophrenic or other delusive disorders; serious organic or metabolic disorders; pregnancy or lactation; and known hypersensitivity to hypericum preparations.	Extract: WS 5572; WS 5573 Dosage: 3X300 mg/day, for 6 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 6 weeks Funding unclear, industry author, provided SJW	Depression Measures: • HAMD, SMD −0.44 (CI −0.85, −0.04) • Responder (HAMD ≥50% decrease), RR 1.47 (CI 0.9, 2.4) Adverse Events: • Placebo: Total AEs 23 out of 49; Infection 2 out of 49; Herpes labialis 2 out of 49; Headache 5 out of 49; Flu-like symptoms 1 out of 49; Cough 1 out of 49; Bronchitis 3 out of 49 • WS 5572: Total AEs 17 out of 49; Infection 0 out of 49; Herpes labialis 0 out of 49; Headache 0 out of 49; Flu-like symptoms 0 out of 49; Cough 0 out of 49; Bronchitis 1 out of 49 • WS 5573: Total AEs 24 out of 49; Infection 1 out of 49; Herpes labialis 0 out of 49; Headache 0 out of 49; Flu-like symptoms 2 out of 49; Cough 2 out of 49; Bronchitis 3 out of 49

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Lecrubier et al., 2002 Country: France Study Design: Multisite RCT, 26 Purpose: to compare the efficacy of H. perforatum extract WS 5570 with that of placebo in a large group of patients suffering from mild to moderate major depressive episode according to DSM-IV Quality Rating: Fair, unclear randomization, adequate blinding, ITT analysis	Number of Participants: 375 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): SJW 40.2 (SD 11.7); Placebo 41.2 (SD 11.4) Gender (% Male): SJW 23.7; Placebo 23.3 Inclusion Criteria: 1) was an outpatient aged 18 to 65 at the time of the screening, 2) provided written, informed consent, 3) had a current major depressive episode of at least 2 weeks' duration that met the criteria of DSM-IV code 296.21, 296.22, 296.31, or 296.32 (mild or moderate depression, single or recurrent episode), and 4) had a total score on the HAMD between 18 and 25 and a score on item 1 ("depressed mood") of 2 or higher at screening and baseline. Exclusion Criteria: Depression of any other type than those specified, any psychiatric disease other than depression, serious suicidal risk (score of 3 or higher on item 3 of the HAMD), or response to placebo during the run-in phase; response was defined as 25% or greater reduction of the HAMD total score.	Extract: WS 5570 Dosage: 300 mg, 3 times a day, for 6 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 6 weeks Industry funding	Depression Measures: • HAMD, SMD −0.26 (CI −0.46, −0.06) • Remission (patients score of 6 or less on HAMD at treatment end), RR 1.56 (CI 1.03, 2.35) • Responder (at least 50% reduction in HAMD total score), RR 1.24 (CI 1, 1.54) Adverse Events: • Placebo: Withdrawals 2 out of 189; Nausea 6 out of 189; Insomnia 2 out of 189; Headache 7 out of 189; Dizziness 4 out of 189; Abdominal pain 4 out of 189 • WS 5570: Withdrawals 2 out of 186; Nausea 9 out of 186; Insomnia 3 out of 186; Headache 3 out of 186; Dizziness 4 out of 186; Abdominal pain 2 out of 186

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Lenoir, Degenring, and Saller, 1999 Country: Switzerland and Germany Study Design: Multisite RCT, 38 Purpose: to investigate the efficacy and tolerability of a new standardized fresh-plant extract obtained from the shoot tips of SJW (hypericum perforatum L.) in the treatment of mild to moderate depression Quality Rating: Poor, unclear blinding, ITT analysis for only tolerability and response rate, efficacy only per protocol, <80% completion	Number of Participants: 348 Diagnosis: MDD-ICD Comorbidities: NA Age (Years): 19–94 (range) Gender (% Male): 26 Inclusion Criteria: mild to moderate depression, aged at least 20 years Exclusion Criteria: a SJW allergy and treatment with antidepressants, tranquilizers, hypnotics, or neuroleptics within the last two weeks immediately prior to the start of the study, and an acute risk of suicide. Any concomitant treatment had to be continued unchanged throughout the 6-week treatment phase.	Extract: Hypericin Dosage: 0.17 mg, 0.33 mg, or 1 mg per day (divided into 3 doses), for 6 weeks Co-interventions: NA Comparator: Other doses of SJW Primary Endpoint: HAMD Power Calculation: No Funding unclear, industry author, provided SJW	Depression Measures: • HAMD-17 relative reduction of 50% observed in all groups; Response rates were 62% in 0.17 mg arm, 65% in 0.33mg arm, and 68% in 1mg arm Adverse Events: • SJW 0.17 mg: Total AEs 17 out of 87; Severe 1 out of 87; Moderate 3 out of 87; Mild 13 out of 87 • SJW 0.33 mg: Total AEs 25 out of 90; Severe 11 out of 90; Moderate 5 out of 90; Mild 9 out of 90 • SJW 1 mg: Total AEs 40 out of 83; Severe 17 out of 83; Moderate 8 out of 83; Mild 15 out of 83

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Liu et al., 2010 Country: China Study Design: RCT Purpose: to assess the effect of SJW on depressive disorder in elderly patients with unstable angina pectoris Quality Rating: Poor, no blinding	Number of Participants: 170 Diagnosis: Rating scale, Other diagnosis, International Society and Federation of Cardiology and World Health Organization criteria Comorbidities: Unstable angina pectoris Age (Years): SJW 67 (SD 2.7); Deanxit 68 (SD 2.8); Psychotherapy 68 (SD 3.0); Control 67 (SD 2.5) Gender (% Male): 50 Inclusion Criteria: The patients aged 65–75 years met the criteria set by International Society and Federation of Cardiology and World Health Organization, and HAMD-17 score ≥17 points. Exclusion Criteria: bipolar disorder, severe mental illness and suicidal tendencies without use of antipsychotropic substances for 3 months before treatment.	Extract: NA Dosage: 300 mg, 3 times a day, for 12 weeks Co-interventions: NA Comparator: Deanxit 10.5 mg per day, cognitive therapy, suggestion therapy, supportive therapy, and rational emotive therapy twice per week; control: oryzanol 20 mg three times per day Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 12 weeks No industry funding	Depression Measures: • HAMD, SMD −0.25 (CI −0.67, 0.17) • Responder (HAMD ≤10 and decrease ≥50%, healing + marked improvement), RR 1.4 (CI 1.04, 1.89) Adverse Events: • Control: Thirsty 0 out of 40; Nausea 0 out of 40; Dizziness 0 out of 40; Constipation 0 out of 40 • Deanxit: Thirsty 0 out of 44; Nausea 0 out of 44; Dizziness 0 out of 44; Constipation 0 out of 44 • Psychotherapy: Thirsty 0 out of 42; Nausea 0 out of 42; Dizziness 0 out of 42; Constipation 0 out of 42 • SJE: Thirsty 3 out of 44; Nausea 2 out of 44; Dizziness 2 out of 44; Constipation 3 out of 44

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Mannel et al., 2010 Country: Germany Study Design: Multisite RCT, 19 Purpose: to focus on the vegetative features in order to test the efficacy of hypericum extract LI 160 prospectively in patients with mild to moderate major depression with atypical characteristics Quality Rating: Good	Number of Participants: 201 Diagnosis: Rating scale, MDD-ICD Comorbidities: NA Age (Years): SJW 47.0 (SD 13.1); Placebo 46.6 (SD 13.8) Gender (% Male): 17 Inclusion Criteria: 18–70 years of age, ICD-10 criteria for mild or moderate depression had to be met, with the adjustment that the duration of symptoms of 3 months was required. A simplified definition of ‘‘atypical depression” was used. The requirement was of a minimum score of two points for at least one of the five items of the HAMD-28 scale, covering the atypical features hypersomnia, increased appetite, and weight gain. Exclusion Criteria: Additionally, a maximum score of one point for items 6 (insomnia late), 12 (somatic symptoms, gastrointestinal), and 16 (loss of weight) of the HAMD-17 scale were allowed, thereby excluding subjects exhibiting vegetative features of melancholic depression. Further, the patients were excluded in case of a history of an episode of melancholic depression, alcohol or substance abuse, organic mental disorders, psychotic disorders, personality disorders, seasonal depression, postpartum depression, and current serious suicidality risk. Patients had to be free of psychotropic drugs for at least 14 days before randomization and of fluoxetine for at least 28 days. No active psychotherapy was permitted before or during the trial. For the actually randomized patients, no placebo washout period was required (all were drug free for at least two weeks before randomization), and eligible patients who had signed written informed consent directly entered the trial. Further exclusion criteria were the use of corticosteroids, including topical, gyrase inhibitors, nor-adrenergic agonists, and magnesium supplements.	Extract: LI 160 Dosage: 300g, 2 times a day, for 8 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD Power Calculation: Yes Follow-Up Time: 8 weeks Industry funding	Depression Measures: • HAMD-17, SMD −0.26 (CI −0.54, 0.02) • Responder, RR 1.2 (CI 0.9, 1.59) Adverse Events: • LI 160: Urinary system disorders 1 out of 100; Total AEs 15 out of 100; Skin and appendage disorders 1 out of 100; Sexual dysfunction 0 out of 100; Serious AEs 0 out of 100; Respiratory system disorders 4 out of 100; Others 2 out of 100; Musculoskeletal system disorders 4 out of 100; Gastrointestinal system disorders 2 out of 100; Central nervous system and peripheral nervous system disorders 1 out of 100 • Placebo: Urinary system disorders 1 out of 100; Total AEs 8 out of 100; Skin and appendage disorders 0 out of 100; Sexual dysfunction 1 out of 100; Serious AEs 1 out of 100; Respiratory system disorders 1 out of 100; Others 4 out of 100; Musculoskeletal system disorders 2 out of 100; Gastrointestinal system disorders 0 out of 100; Central nervous system and peripheral nervous system disorders 0 out of 100

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Montgomery, Hübner, and Grigoleit, 2000 Country: United Kingdom Study Design: Multisite RCT, 18 Purpose: the efficacy and tolerability of SJW (Hypercium) extract LI 160 was compared with placebo in patients with a mild to moderate depressive disorder Quality Rating: Poor, not enough information	Number of Participants: 248 Diagnosis: MDD-DSM Comorbidities: NA Age (Years): NA Gender (% Male): NA Inclusion Criteria: defined by the DSM-IV Exclusion Criteria: NA	Extract: LI 160 Dosage: 300mg, 3 times a day, for 12 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 6 weeks Funding unclear, NR	Depression Measures: • Responder HAMD, RR 1.46 (CI 1.159, 1.85) Adverse Events: • LI 160: Drug-related serious side effects 0 out of 124 • Placebo: Drug-related serious side effects 0 out of 124

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Moreno et al., 2005 Country: Brazil Study Design: RCT Purpose: assessed the efficacy and safety of hypericum perforatum in comparison with fluoxetine, in an 8-week double-blind trial in patients with mild to moderate depression Quality Rating: Fair, ITT analysis but randomization and blinding unclear	Number of Participants: 66 Diagnosis: Rating scale Comorbidities: NA Age (Years): 40.5 (SD 10.7) Gender (% Male): 17 Inclusion Criteria: baseline score of at least 10 points in the HAMD-21 and a maximum baseline score of 24 points. Exclusion Criteria: Patients with other types of depression, psychosis, personality disorders (such as borderline or depressive), bipolar disorders, suicidal ideation, uncontrolled organic disease, or history of alcohol or drug abuse 1 year prior to the screening; patients who had abnormal laboratorial tests or a history of seizures; and patients who had been treated with electroconvulsotherapy or had taken any investigational drug up to 30 days before screening were excluded. Patients who used MAO-inhibitors 2 weeks prior to the screening, other antidepressants, or any other drug (except benzodiazepines in doses equivalent to diazepam 10 mg/day by mouth 1 week prior to the screening), and those who had already been treated with fluoxetine were also excluded.	Extract: NA Dosage: 300 mg, 3 times a day, for 8 weeks Co-interventions: NA Comparator: Fluoxetine, placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 8 weeks Industry funding	Depression Measures: • Remission (total HAMD score ≤7), RR 0.22 (CI 0.05, 0.86) • Remission (total HAMD score ≤7), RR 0.289 (CI 7.000, 1.21) • Responder (50% decrease in HAMD or MADRS total scores), RR 0.36 (CI 0.140, 0.95) • Responder (50% decrease in HAMD or MADRS total scores), RR 0.47 (CI 0.18, 1.27) Adverse Events: • No differences between the 3 groups regarding safety measures, including vital signs. Tension, nausea, postural dizziness, menorrhagia, and diminished sexual desire were more frequent in the fluoxetine group at week 4. • In the 8th week, there was a higher incidence of insomnia, headache, and diarrhea in the fluoxetine group.

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Pakseresht et al., 2012 Country: Iran Study Design: RCT Purpose: to assess the effect of hypericum perforatom (perforan), in combination with tricyclic antidepressants in MDD treatment Quality Rating: Fair, unclear double-blinding but stated double-blind and all completed	Number of Participants: 40 Diagnosis: Rating scale, Other diagnosis, diagnosed depression, method unspecified Comorbidities: NA Age (Years): SJW 29.8 (SD 6.2); Placebo 30 (SD 16.6) Gender (% Male): SJW 50; Placebo 45 Inclusion Criteria: diagnosed with mild or moderate MDD for six weeks, between 18–55 years of age. Optioned Beck Depression Inventory (BDI) was performed before treatment and only the patients who earned 16–46 points were taken into account in the study. Exclusion Criteria: pregnancy and lactation, the presence of clinically significant organic or neurological disorders, Axis II disorder, comorbid disorder in Axis I, consumption of alcohol and other addictive substances except nicotine and caffeine, symptoms that caused or worsen psychotic depression symptoms, symptoms that required hospitalization and emergency action, patients with suicidal thoughts, history of receiving electroconvulsive therapy in the last 3 months, any allergies to medicine (particularly hypericum perforatom), or taking lithium, anticonvulsants, sumatriptan, L. dopa, SSRI, buspirone, ergot compounds, selegiline, stimulants, anti-congestive medications, contraceptives, cimetidine, theophylline, or thyroid hormones. At the beginning of the study, blood tests were done for blood sugar, fat, and liver and kidney function; patients were excluded if test results were not normal.	Extract: NA Dosage: 300 mcg, 3 times a day, for 6 weeks Co-interventions: Nortriptyline 75–100 mg daily, imipramine and amitriptyline 100–150 mg daily, for 6 weeks Comparator: Nortriptyline 75–100 mg, imipramine, amitriptyline 100–150 mg daily, Placebo Primary Endpoint: BDI Power Calculation: No Follow-Up Time: 6 weeks No industry funding	Depression Measures: • BDI, SMD −7.00 (CI −0.69, 0.550) Adverse events: • No sexual side effects. • Hypericum group: 3 females developed photosensitivity. Lower gastrointestinal complications than those taking placebo.

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Philipp, Kohnen, and Hiller, 1999 Country: Germany Study Design: Multisite RCT, 18 Purpose: to assess the efficacy and safety of hypericum extract (STEI 300, Steiner Arzneimittel, Berlin) compared with imipramine and placebo inpatients in primary care with a current episode of moderate depression Quality Rating: Fair, unclear randomization, adequate double-blinding, ITT analysis	Number of Participants: 263 Diagnosis: Rating scale, MDD-ICD Comorbidities: NA Age (Years): 47 (SD 12) Gender (% Male): 25 Inclusion Criteria: Men and women aged 18–65; diagnosis of a moderate depressive episode according to ICD-10 codes F32.1 and F33.1; minimum total score of 18 on the 17-item version of the HAMD; a CGI rating of severity (item 1) of moderately, markedly, or severely ill; depression duration a minimum of four weeks and a maximum of two years. Exclusion Criteria: Mild and severe depressive disorders according to ICD-10 codes F32.0, F33.0, F32.2, F33.2, F32.3, and F33.3; bipolar disorders according to ICD-10 codes F31.x; comorbidity from alcohol or drug dependence according to ICD-10 codes F10–F19; suicidal risk (assessed by item 10 of the MADRS); long-term prophylaxis with lithium or carbamazepine; non-sufficient washout phase of previous psychotropic drug; any interfering psychotropic drug taken concurrently; any previous long-term (>3 months) treatment with benzodiazepines; patients at general and specific risk (imipramine contraindications).	Extract: STEI 300 Dosage: 350 mg, 3 times a day, for 8 weeks Co-interventions: NA Comparator: Imipramine, placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 8 weeks Industry funding	Depression Measures: • HAMD, SMD −0.16 (CI −0.42, 0.11) • HAMD, SMD −0.42 (CI −0.76, −7.00) • Responder (≥50% improvement in total HAMD score), RR 1.07 (CI 0.88, 1.3) • Responder (≥50% improvement in total HAMD score), RR 1.43 (CI 1.03, 2) Quality of Life: • Quality of life SF-36 mental component, SMD 0.11 (CI −0.15, 0.38) • Quality of life SF-36 mental component, SMD 0.46 (CI 0.11, 0.81) • Quality of life SF-36 physical component, SMD 0.23 (CI −0.04, 0.5) • Quality of life SF-36 physical component, SMD 0.35 (CI 0.01, 0.7) Adverse Events: • Imipramine: Withdrawal 1 out of 110; Serious AEs 0 out of 110; Palpitation 6 out of 110; Nausea 12 out of 110; Headache 6 out of 110; Dry mouth 42 out of 110; Dizziness 7 out of 110; Constipation 7 out of 110 • Placebo: Withdrawal 0 out of 47; Serious AEs 1 out of 47; Palpitation 0 out of 47; Nausea 1 out of 47; Headache 1 out of 47; Dry mouth 6 out of 47; Dizziness 1 out of 47; Constipation 3 out of 47 • SJW Extract: Withdrawal 0 out of 106; Serious AEs 0 out of 106; Palpitation 4 out of 106; Nausea 8 out of 106; Headache 3 out of 106; Dry mouth 7 out of 106; Dizziness 1 out of 106; Constipation 4 out of 106

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Rahman et al., 2008 Country: Pakistan Study Design: RCT Purpose: to compare improvement in symptoms of mild to moderate depression after treatment with hypericum perforatum (SJW extract) and placebo Quality Rating: Poor, <80% follow up, no ITT analysis	Number of Participants: 225 Diagnosis: Rating scale, MDD-ICD Comorbidities: NA Age (Years): SJW 33.89 (SD 10.884); Placebo 36.29 (SD 12.478) Gender (% Male): SJW 23.2; Placebo 21.4 Inclusion Criteria: Patients of both sexes, between the ages of 18 to 65, with no associated physical disease and who gave their consent were recruited for the study. Mild to moderate depression was assessed according to ICD-10; F32.0 Mild depressive episode or F33.0 Recurrent depressive disorder, current episode mild and F32.1 Moderate depressive episode or F33.1 Recurrent depressive disorder, current episode moderate. Participants were required to have a total score between the ranges of 15–22 on 17-item HAMD. Exclusion Criteria: Any patients with depression secondary to organic illness and atypical cases that may carry different diagnosis were not included. Patients belonging outside of Karachi were also not included because of inherent difficultly in follow-up.	Extract: NA Dosage: 300 mg, 3 times a day, for 6 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 6 weeks Industry funding	Depression Measures: • HAMD score decreased by about 51% in SJW arm and about 46% in placebo, after 6 weeks. Adverse Events: • Placebo: Withdrawal 2 out of 56; Photosensitivity 4 out of 56; Palpitations 0 out of 56; Nausea 8 out of 56; Headache 16 out of 56; Dry mouth 0 out of 56; Dizziness 4 out of 56 • SJW: Withdrawal 0 out of 56; Photosensitivity 4 out of 56; Palpitations 4 out of 56; Nausea 16 out of 56; Headache 16 out of 56; Dry mouth 8 out of 56; Dizziness 12 out of 56

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Schrader, Meier, and Brattström, 1998 Country: Germany Study Design: Multisite RCT, 16 Purpose: compares the efficacy and tolerability of hypericum administered as a concentrated ethanolic extract of SJW (Ze 117) with patients with mild-moderate depression (ICD-10; F32.0 mild; F32.1 moderate). Quality Rating: Good	Number of Participants: 162 Diagnosis: Rating scale, MDD-ICD Comorbidities: NA Age (Years): SJW 47 (SD 32–59.25, 25–75% range); Placebo 39 (SD 30–59.25, 25–75% range) Gender (% Male): SJW 28; Placebo 38 Inclusion Criteria: over the age of 18 years presenting with mild-moderate depression defined according to ICD-10 (F32.0; F32.1) and who had total scores between 16 and 24 on the HAMD were admitted to the study. Exclusion Criteria: Excluded from entry were those who had taken part in other clinical trials in the previous 4 weeks or during the study itself, those suffering from psychiatric disorders that might impair accurate history, patients unable or unwilling to give written informed consent, presence of neoplasia, Parkinson's or Alzheimer's disease, pregnancy or inadequate contraception, risk of suicide (score ≥2 on suicidality item of HAMD), known hypersensitivity to SJW, severe concomitant systemic diseases, chronic alcohol or drug abuse, and concomitant psychotherapy or drug therapy that could influence the assessment of efficacy variables.	Extract: Ze 117 Dosage: 250 mg, 2 times a day, for 6 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 6 weeks Funding unclear, industry author, provided SJW	Depression Measures: • HAMD 21-item, SMD −1.22 (CI −1.56, −0.88) • Remission (improvement of more than 4 points HAMD), RR 3.43 (CI 2.29, 5.14) • Responder (improvement of ≥50% HAMD or total score ≤10), RR 3.7 (CI 2.12, 6.46) Adverse Events: • Placebo: Total AEs 5 out of 81; Syncope - severe 1 out of 81; Serious AEs 1 out of 81; Paraesthesia - mild 1 out of 81; Melancholia - moderate 0 out of 81; Dry mouth - mild 0 out of 81; Diarrhea - moderate 0 out of 81; Acute deterioration - moderate 0 out of 81; Abdominal pain - moderate 3 out of 81 • SJW: Total AEs 6 out of 81; Syncope - severe 0 out of 81; Serious AEs 0 out of 81; Paraesthesia - mild 0 out of 81; Melancholia - moderate 1 out of 81; Dry mouth - mild 1 out of 81; Diarrhea - moderate 1 out of 81; Acute deterioration - moderate 1 out of 81; Abdominal pain - moderate 2 out of 81

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Schrader, 2000 Country: Germany Study Design: Multisite RCT, 7 Purpose: to evaluate the clinical efficacy of hypericum (SJW) extract tablets (Ze 117 ethanol extract 50% weight/weight, drug-extract ratio 4-7:1) against one of the most widely used SSRIs, fluoxetine (Prozac®), using effective dosages as recommended by the manufacturers; both were given for a period of 6 weeks Quality Rating: Fair, unclear randomization, adequate blinding, ITT analysis	Number of Participants: 240 Diagnosis: Rating scale, MDD-ICD Comorbidities: NA Age (Years): SJW 46 (SD 19); Fluoxetine 47 (SD 17) Gender (% Male): SJW 29; Fluoxetine 41 Inclusion Criteria: Subjects of both genders, aged 18 years or older, gave their written informed consent prior to enrollment, had a baseline depression score (21-item HAMD) of 16–24, inclusive, and fulfilled the diagnostic criteria for mild-moderate depression. Exclusion Criteria: Excluded from entry were those with a history of alcohol/substance abuse or dependence, dementia, or other severe intellectual impairment that might preclude informed consent; a history of seizures; glaucoma; pituitary deficiency; suicidal ideation (score 2–4 on HAMD item 3); thyroid or parathyroid pathology; Parkinson's disease; or any serious concomitant medical condition. Also excluded were pregnant or breastfeeding women. Patients previously treated with MAO inhibitors underwent a washout period of 2 weeks, and this was extended to 5 weeks for those previously receiving SSRIs. The following concomitant treatments were not allowed during the study: quinidine, anticholinergic drugs, cimetidine, cardiac glycosides neuroleptics, sympathomimetic drugs, MAO inhibitors, tryptophan, and any other antidepressant.	Extract: Ze 117 Dosage: 250 mg, 2 times a day, for 6 weeks Co-interventions: NA Comparator: Fluoxetine Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 6 weeks Industry funding	Depression Measures: • HAMD, SMD 0.15 (CI −0.11, 0.4) • Responder (≥50% decrease in HAMD or final score of ≤10), RR 1.51 (CI 1.149, 1.97) Adverse Events: • Fluoxetine: Withdrawals due to AE 1 out of 114; Total AEs 38 out of 62; Patients reporting AEs possibly/probably related to drug 26 out of 114; All AEs possible/probably related to drug 34 out of 47 • SJW: Withdrawals due to AE 0 out of 125; Total AEs 24 out of 62; Patients reporting AEs possibly/probably related to drug 10 out of 125; All AEs possible/probably related to drug 13 out of 47

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Shelton et al., 2001 Country: United States Study Design: Multisite RCT, 11 Purpose: to compare the efficacy and safety of a standardized extract of SJW with placebo in outpatients with major depression Quality Rating: Good, achieved adequate randomization and double-blinding, ITT analysis, >80% follow-up, valid measures, appropriate attention to confounders	Number of Participants: 200 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): SJW 41.4 (SD 12.5); Placebo 43.3 (SD 13.7) Gender (% Male): SJW 35.1; Placebo 37.2 Inclusion Criteria: Physically healthy male or female outpatients, 18 years or older, diagnosed as having major depressive disorder, single episode or recurrent, without psychotic features according to the DSM-IV, of at least 4 weeks' duration. Participants had a score of at least 20 on the HAMD (17-item scale) at baseline. Exclusion Criteria: Current cognitive disorder, posttraumatic stress disorder, eating disorder, or a substance use disorder in the last 6 months; panic disorder in the last year; or current or past history of bipolar disorder or any psychotic disorder, or borderline, antisocial, or schizotypal personality disorder. Anyone with a prior adequate trial of SJW (at least 450 mg/d) for the treatment of depression or those who had taken SJW for any reason in the last month were excluded. To reduce the potential for including a treatment non-responsive sample, participants who had failed to respond to a trial of an antidepressant (fluoxetine hydrochloride, 20 mg/d, for at least 4 weeks or the equivalent) in the current episode or who had failed to respond to more than 1 adequate trial of antidepressant in a previous episode were also excluded. Patients could not take other psychotropic medications during study participation, with the exception of zolpidem tartrate, which was allowed up to 10 mg/d for sleep for the first 3 weeks of the trial. All participants received a physical examination, electrocardiogram, hematological and blood chemistry screening, and urine testing for illicit drugs. Persons in psychotherapy were allowed if they were in therapy for at least 3 months prior to baseline, and if the frequency of sessions did not change during participation. Women also received a urine pregnancy test.	Extract: NA Dosage: 300 mg a day, for 8 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD Power Calculation: Insufficient power (posthoc analysis) Follow-Up Time: 8 weeks Funding Unrestricted grant/industry funding. but no conflict	Depression Measures: • BDI, SMD −0.28 (CI −0.56, 0) • Remission (HAMD <8 or <7), RR 0.15 (CI 0.09, 0.24) • Responder HAMD, RR 1.42 (CI 0.84, 2.4) Adverse Events: • Placebo: Withdrawal due to AEs 1 out of 102; Headaches 25 out of 100 • SJW: Withdrawal due to AEs 1 out of 98; Headaches 39 out of 95

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Szegedi et al., 2005 Country: Germany Study Design: Multisite RCT, 21 Purpose: to investigate the efficacy of hypericum extract WS 5570 (SJW) compared with paroxetine in patients with moderate to severe major depression Quality Rating: Fair, adequate randomization, blinding, <80% , ITT analysis	Number of Participants: 251 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): SJW 49.0 (SD 11.0); Paroxetine 45.5 (SD 11.5) Gender (% Male): SJW 30; Paroxetine 32 Inclusion Criteria: All participants were 18–70 years old and had single or recurrent moderate or severe episodes of unipolar major depression without psychotic features (DSM-IV) 296.22, 296.23, 296.32, 296.33) persisting for two weeks to a year. At screening and baseline, all participants had to have a total score of ≥22 points on the 17-item HAMD and ≥2 points for the item "depressive mood." The diagnosis of depression was based on the mini-international neuropsychiatric interview. There were no restrictions regarding ethnic group. Exclusion Criteria: We excluded anyone with a decrease in total depression score of ≥25% during the run-in, or with a diagnosis of schizophrenia, acute anxiety disorder, adjustment disorder, depressive disorder of any type not stated above, bipolar disorder, organic mental disorder, acute posttraumatic stress disorder, or substance abuse disorder; increased risk of suicide (defined by a score ≥4 for item 10 of the MADRS), who had previously attempted suicide, or who had not responded to more than one adequate treatment (equivalent to 150 mg/day amitriptyline for 6 weeks) in the present episode. Participants were not allowed to take other psychotropic medication or psychotherapy during the study (in case of previous antidepressant medication, an appropriate washout period of five half lives had to be observed).	Extract: WS 5570 Dosage: 300–600 mg, 3 time a day, for 6 weeks Co-interventions: NA Comparator: Paroxetine Primary Endpoint: HAMD Power Calculation: Yes Follow-Up Time: 6 weeks Industry funding	Depression Measures: • HAMD, SMD −0.34 (CI −0.6, −0.09) • Remission, RR 1.42 (CI 1.05, 1.91) • Responder HAMD, RR 1.18 (CI 0.98, 1.42) Adverse Events: • Paroxetine: Upper abdominal pain 9 out of 126; Total AEs 269 out of 126; Sleep disorder 10 out of 126; Serious AEs 0 out of 126; Nausea 21 out of 126; Increased sweating 13 out of 126; Headache 14 out of 126; Fatigue 16 out of 126; Dry mouth 35 out of 126; Dizziness 24 out of 126; Diarrhea 23 out of 126 • SJW: Upper abdominal pain 12 out of 125; Total AEs 172 out of 125; Sleep disorder 5 out of 125; Serious AEs 2 out of 125; Nausea 9 out of 125; Increased sweating 9 out of 125; Headache 13 out of 125; Fatigue 14 out of 125; Dry mouth 16 out of 125; Dizziness 9 out of 125; Diarrhea 12 out of 125

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Uebelhack et al., 2004 Country: Germany Study Design: RCT Purpose: to compare the clinical efficacy and tolerability of oral hypericum extract STW 3-VI (Laif®) 900 mg once daily with that of placebo Quality Rating: Good	Number of Participants: 140 Diagnosis: MDD-DSM, Rating scale, MDD-ICD Comorbidities: NA Age (Years): SJW 46.4 (SD 12.5); Placebo 43.3 (SD 12.6) Gender (% Male): SJW 30; Placebo 36 Inclusion Criteria: diagnosis of moderate depressive disorder according to ICD-10 F32.1 or F33.1 and DSM-IV; total HAMD-17 score of 20 to 24 at the first and second examination; aged 18 to 70 years; and contraception use by women of childbearing age. All patients were informed of the importance, aim, and procedure of the study before entry and gave written consent for their participation. Exclusion Criteria: depression resistant to treatment; a known history of schizophrenic, psychotic, epileptic or dementia disorders; depression caused by another severe disease; known intolerance of the study medication; known photosensitivity; specific psychotherapy during the study and during the last 2 months before study entry; use of psychotropic drugs (e.g., antidepressant, neuroleptic, and anxiolytic agents) during the study and during the last 6 weeks before study entry; concomitant use of coumarin anticoagulants; or known history of attempted suicide or acute suicidality (item 3 of the HAMD-17 >2). Additional exclusion criteria were participation in a clinical trial within the last 30 days, simultaneous participation in another clinical trial, or attendance in this trial at an earlier time; existence of psychiatric disorders that could influence the results of the study; epilepsy; a personal or family history of melanoma; pregnancy or lactation; chronic alcohol or drug dependency; HIV infection or a diagnosis of AIDS or a neoplastic disease; or clinically relevant deviations from normal laboratory values due to severe forms of other illnesses. Concomitant medications that would not influence the results were allowed. Changes in concomitant medications during the study period (e.g., due to adverse events) were questioned at each examination and documented on the clinical report form.	Extract: STW 3-VI Dosage: 900 mg a day, for 6 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 6 weeks Funding unclear, NR	Depression Measures: • HAMD, SMD −1.79 (CI −2.18, −1.4) • Remission (no need for any further medication), RR 7.5 (CI 2.79, 20.17) • Responder (HAMD decrease ≥50% or total score less than 10 points), RR 10.25 (CI 3.88, 27.09) Adverse Events: • Placebo: Total AEs 7 out of 70; Moderate AEs 1 out of 70; Mild AEs 6 out of 70; Gastrointestinal symptoms possibly/probably related to study medication 0 out of 70 • SJW: Total AEs 16 out of 70; Moderate AEs 4 out of 70; Mild AEs 12 out of 70; Gastrointestinal symptoms possibly/probably related to study medication 2 out of 70

Study Details	Participants	Intervention/Treatment	Outcomes/Results
van Gurp et al., 2002 Country: Canada Study Design: RCT Purpose: to compare the change in severity of depressive symptoms and occurrence of side effects in primary care patients treated with SJW and sertraline Quality Rating: Fair, adequate randomization and blinding, ITT analysis, <80% completed	Number of Participants: 90 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): SJW 40.9 (SD 11.6); Sertraline 39.1 (SD 10.2) Gender (% Male): SJW 36.4; Sertraline 41.5 Inclusion Criteria: People aged 18 to 65 years fluent in French or English were eligible to participate if they had been diagnosed with major depression using DSM-IV criteria and had a HAMD score of ≥16. Exclusion Criteria: pregnant, lactating, not using acceptable contraception, or at serious risk of suicide; had other indications for hospitalization (including delusions or hallucinations); or had a history of drug or alcohol abuse in the previous 3 months, other DSM-IV comorbid conditions, or serious medical illnesses. Patients who had concomitantly used other psychoactive drugs regularly during the previous 2 weeks (4 weeks if taking fluoxetine), with the exception of bedtime sedative-anxiolytics	Extract: NR Dosage: 1-2 300 mg, 3 times a day, for 12 weeks Co-interventions: NA Comparator: Sertraline Primary Endpoint: HAMD Power Calculation: Yes Follow-Up Time: 12 weeks Funding Unrestricted grant/industry funding but no conflict	Depression Measures: • HAMD, SMD −0.25 (CI −0.67, 0.17) Adverse Events: • Sertraline: Urinary problems 7 out of 34; Tremor 5 out of 34; Sweating 13 out of 34; Sleep disturbance 24 out of 34; Sexual difficulties 15 out of 34; Serious AEs 1 out of 34; Pain 8 out of 34; Nausea or vomiting 17 out of 34; Muscle spasms 5 out of 34; Muscle or joint stiffness 12 out of 34; Lack of appetite 11 out of 34; Heart palpitations 7 out of 34; Headaches 14 out of 34; Fatigue 21 out of 34; Dry mouth 20 out of 34; Dizziness 11 out of 34; Difficulty digesting 14 out of 34; Diarrhea 17 out of 34; Blurred vision 7 out of 34; Anxiety 18 out of 34 • SJW: Urinary problems 7 out of 44; Tremor 8 out of 44; Sweating 7 out of 44; Sleep disturbance 23 out of 44; Sexual difficulties 5 out of 44; Serious AEs 0 out of 44; Pain 5 out of 44; Nausea or vomiting 4 out of 44; Muscle spasms 5 out of 44; Muscle or joint stiffness 8 out of 44; Lack of appetite 10 out of 44; Heart palpitations 4 out of 44; Headaches 18 out of 44; Fatigue 19 out of 44; Dry mouth 16 out of 44; Dizziness 5 out of 44; Difficulty digesting 8 out of 44; Diarrhea 10 out of 44; Blurred vision 6 out of 44; Anxiety 18 out of 44

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Volz, Eberhardt, and Grill, 2000 Country: Germany Study Design: Multisite RCT, 17 Purpose: investigate the extract D-0496 versus placebo in 140 patients suffering from a mild or moderate depressive episode according to DSM-IV, treated in a double-blind manner with one capsule in the morning and one capsule in the evening Quality Rating: Good	Number of Participants: 140 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): 47 Gender (% Male): 19 Inclusion Criteria: Mild to moderate major depression (DSM-IV); between 18 and 65 years old; duration of depressive episode between 2 weeks and 6 months; HAMD ≥18; Exclusion Criteria: Suicidal tendencies; severe depression; improvement for more than 4 points on HAMD during run-in phase; psychotic episodes; further DSM-IV diagnoses; relevant somatic diseases; hypersensitivity to SJW; pregnant or currently breastfeeding; current use of psychopharmaceuticals or psychotherapy.	Extract: D-0496 (hypericin) Dosage: 250 mg, 2 times a day, for 6 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: HAMD Power Calculation: Yes Follow-Up Time: 6–8 weeks Funding unclear, industry author, provided SJW	Depression Measures: • HAMD, SMD −0.41 (CI −0.75, −0.08) • Responder CGI (at least much improved), RR 1.35 (CI 1.01, 1.82) Adverse Events: • Placebo: skin changes 2 out of 19; respiratory infection 3 out of 19; other 5 out of 19; gastrointestinal event 9 out of 19; any AE 22 out of 19; urinary tract infection 3 out of 19 • SJW: skin changes 3 out of 12; respiratory infection 7 out of 12; other 5 out of 12; gastrointestinal event 2 out of 12; any AE 18 out of 12; urinary tract infection 1 out of 12

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Vorbach, Arnoldt, and Hubner, 1997 Country: Germany Study Design: Multisite RCT, 20 Purpose: to compare 1800 mg LI 160/die to 150 mg imipramine/die in severely depressed patients according to ICD-10 Quality Rating: Poor, no ITT analysis	Number of Participants: 209 Diagnosis: MDD-ICD Comorbidities: NA Age (Years): SJW 48.8 (SD 12.0); Imipramine 50.1 (SD 11.8) Gender (% Male): SJW 27; Imipramine 25 Inclusion Criteria: Men and women aged 18 to 70, ICD-10 F332 (severe episode of a major depressive disorder, recurrent, without psychotic symptoms). At least two prior episodes of at least 2 weeks duration were obligatory. Exclusion Criteria: Patients with a suicidal tendency, hallucinations, and depressive delusional content. Equally, patients with possible pre-existing schizophrenic disorders or pronounced agitation, chronic alcohol or drug dependency, and acute confusional states. The patients were not allowed to take any psychotropic medication besides the investigational drugs, with the exception of chloral hydrate in the case of sleep disturbances. Lithium was allowed if it had been prescribed at least 3 months before the trial and was continued with an unchanged daily dose. If patients had been pretreated with MAO-inhibitors, this regimen had to be discontinued at least 14 days before the start of the trial.	Extract: LI 160 Dosage: 3 x 600 mg a day, for 6 weeks Co-interventions: NA Comparator: imipramine Primary Endpoint: HAMD Power Calculation: No Follow-Up Time: 6 weeks Funding unclear, industry author, provided SJW	Depression Measures: • Responder (reduction of ≥ 0% HAMD), RR 0.86 (CI 0.61, 1.22) • Total HAMD, SMD 0.17 (CI −0.11, 0.44) Adverse Events: • Imipramine: Tremor 4 out of 102; Total AEs 83 out of 102; Tiredness/sedation 8 out of 102; Sweating 8 out of 102; Sleep disorders 2 out of 102; Restlessness 6 out of 102; Pressure in the head 3 out of 102; Palpitations 3 out of 102; Gastric symptoms 9 out of 102; Dry mouth 16 out of 102; Dizziness 2 out of 102; Constipation 5 out of 102; Allergic skin reactions 2 out of 102 • LI 160: Tremor 2 out of 107; Total AEs 37 out of 107; Tiredness/sedation 5 out of 107; Sweating 0 out of 107; Sleep disorders 0 out of 107; Restlessness 6 out of 107; Pressure in the head 0 out of 107; Palpitations 0 out of 107; Gastric symptoms 5 out of 107; Dry mouth 3 out of 107; Dizziness 5 out of 107; Constipation 0 out of 107; Allergic skin reactions 1 out of 107

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Wheatley, 1997 Country: United Kingdom Study Design: Multisite RCT, 19 Purpose: LI 160 (total daily dose: 900 mg) was compared with the sedating tricyclic amitriptyline (total daily dose: 75 mg) to treat mild and moderate depression Quality Rating: Good	Number of Participants: 165 Diagnosis: MDD-DSM, Rating scale Comorbidities: NA Age (Years): SJW 42 (range: 20–64); Amitriptyline 38 (range: 24–65) Gender (% Male): SJW 15.7; Amitriptyline 23.3 Inclusion Criteria: Age between 20 and 65 years, a current major depressive episode according to DSM-IV criteria, and an initial HAMD (17-item form) score between 17 and 24. Exclusion Criteria: Pregnancy or lactation, known history or presence of serious renal, hepatic, or cardiovascular diseases, blood dyscrasia or anaemia, organic brain diseases, and the established exclusion criteria for use of tricyclic antidepressants. Risk of suicide and/or a HAMD score of ≥3 on item 3 (suicidality) was also not allowed. The use of other psychoactive medication with the exception of temazepam (10–20 mg/day), zopiclone (7.5 mg/day), or zoplidem (5–10 mg/day) as hypnotics was contraindicated. Antidepressants had to be omitted at least 14 days before the placebo run-in period; in the case of fluoxetine, 42 days were required. Patients who improved during the placebo run-in phase to a HAMD total score of <16 or with a reduction of >25% were also excluded.	Extract: LI 160 Dosage: 300 mg, 3 times a day, for 6 weeks Co-interventions: NA Comparator: Amitriptyline Primary Endpoint: HAMD Power Calculation: Yes Follow-Up Time: 6 weeks Funding unclear, NR	Depression Measures: • HAMD, SMD 0.74 (CI 0.42, 1.07) • Responder (HAMD total score <10 or ≥50% decrease), RR 0.77 (CI 0.62, 0.95) Adverse Events: • Amitriptyline: Pruritus 1 out of 78; Headache 2 out of 78; Constipation 1 out of 78; Sleepiness 8 out of 78; Nausea/vomiting 6 out of 78; Lethargy 3 out of 78; Dry mouth 32 out of 78; Drowsiness 11 out of 78; Dizziness 6 out of 78 • LI 160: Pruritus 2 out of 87; Headache 6 out of 87; Constipation 4 out of 87; Sleepiness 2 out of 87; Nausea/vomiting 6 out of 87; Lethargy 1 out of 87; Dry mouth 4 out of 87; Drowsiness 1 out of 87; Dizziness 1 out of 87

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Witte et al., 1995 Country: Germany Study Design: Multisite RCT, 5 Purpose: to provide evidence for the tolerability and effectiveness of a new highly concentrated SJW extract, psychotonin forte Quality Rating: Good	Number of Participants: 97 Diagnosis: Rating scale, MDD-ICD Comorbidities: NA Age (Years): 44.7 (SD 10.9) SJW; 41.6 (SD 12.5) Placebo Gender (% Male): 31 (SJW); 37 (Placebo) Inclusion Criteria: ICD-10 defined depression with a HAMD score of 16 or more. Exclusion Criteria: Other psychopharmaceutical usage in 4 weeks before the study began; psychiatric diseases other than those defined ICD-10 F32.1; chronic depression; suicide attempts; known adverse reactions to SJW; known sun allergy; renal deficiency; acute or chronic liver disease; alcohol, medication, or drug dependency; or trying to conceive.	Extract: Psychotonin forte Dosage: 100–120 mg, 2 times a day, for 6 weeks Co-interventions: NA Comparator: Placebo Primary Endpoint: At least 50% reduction in HAMD score, or score less than 10 on HAMD Power Calculation: No Follow-Up Time: 6 weeks Funding unclear, NR	Depression Measures: • HAMD responder, RR 1.44 (CI 1.07, 1.92) • Remission (not at all ill, physician assessment), RR 4.08 (CI 1.83, 9.1) Adverse Events: • Placebo: stomach pressure 1 out of 33; not well tolerated 4 out of 33 • SJW: stomach pressure 0 out of 39; not well tolerated 3 out of 39

Study Details	Participants	Intervention/Treatment	Outcomes/Results
Woelk, 2000 Country: Germany Study Design: Multisite RCT, 40 Purpose: to compare the efficacy and tolerability of hypericum perforatum (SJW extract) with imipramine in patients with mild to moderate depression Quality Rating: Fair, double-blinded RCT, well described intervention/ outcomes, ITT analysis, <80% follow-up in imipramine group	Number of Participants: 324 Diagnosis: Rating scale, MDD-ICD Comorbidities: NA Age (Years): SJW 46.5 (SD 12.7); Imipramine 45.4 (SD 12.8) Gender (% Male): SJW 29; Imipramine 29 Inclusion Criteria: Men and women aged 18 or older, with mild to moderate depression without increased suicidal ideation, if they fulfilled ICD-10 criteria for a depressive episode or recurrent depressive disorder (ICD-10 codes F32.0 or F33.0 and F32.1 or F33.1). Score >18 on the 17-item HAMD on two consecutive visits. All participants gave written, informed consent before entering the study. Exclusion Criteria: pregnant or breast feeding, premenopausal and not using contraception, known to be allergic to the drugs being studied, or had a serious disease that in the investigator's opinion should preclude their entry to the study. They were also excluded if they had abnormal thyroid function or other relevant abnormalities on laboratory testing, or if they had bipolar disorder, previous serious psychiatric disease, or misused alcohol or drugs. Participants who had taken any of the following medications within the past 14 days were also excluded: MAO inhibitors, antidepressant drugs, lithium, antipsychotic drugs, neuroleptic drugs, cimetidine, oral corticosteroids, anticonvulsants, theophylline, or thyroid hormones. Owing to the 50% chance of receiving imipramine in the study, benzodiazepines were allowed at a maximum daily dose of 10 mg diazepam for not longer than three consecutive days on not more than three occasions over the six weeks of the study.	Extract: Ze 117 Dosage: 250 mg, 2 times a day, for 6 weeks Co-interventions: NA Comparator: imipramine Primary Endpoint: HAMD Power Calculation: Yes Follow-Up Time: 6 weeks Funding Unrestricted grant/industry funding but no conflict	Depression Measures: • HAMD responder (≥50% decrease in HAMD), RR 1.08 (CI 0.83, 1.4) Adverse Events: • Imipramine: withdrawals caused by AE 26 out of 167; Total AEs 238 out of 167; Sweating 13 out of 167; Nausea 12 out of 167; Headache 6 out of 167; Dry mouth 41 out of 167; Dizziness 12 out of 167; Asthenia 11 out of 167; AE possible/probably related to drug treatment 125 out of 167 • SJW: withdrawals caused by AE 4 out of 157; Total AEs 121 out of 157; Sweating 2 out of 157; Nausea 1 out of 157; Headache 3 out of 157; Dry mouth 13 out of 157; Dizziness 0 out of 157; Asthenia 2 out of 157; AE possible/probably related to drug treatment 50 out of 157

NR = not reported.
N/A = not available.
SD = standard deviation.

References

Balshem, H., M. Helfand, H. J. Schünemann, A. D. Oxman, R. Kunz, J. Brozek, G. E. Vist, Y. Falck-Ytter, J. Meerpohl, S. Norris, and G. H. Guyatt, "GRADE Guidelines: 3. Rating the Quality of Evidence," *Journal of Clinical Epidemiology,* Vol. 64, No. 4, 2011, pp. 401–406. PMID: 21208779

Behnke, K., G. S. Jensen, H. J. Graubaum, and J. Gruenwald, "Hypericum Perforatum Versus Fluoxetine in the Treatment of Mild to Moderate Depression," *Advances in Therapy,* Vol. 19, No. 1, January–February 2002, pp. 43–52. PMID: 12008860

Bernhardt, M., E. Liske, and L. Ebeling, "Hypericum Perforatum in der Therapie Leichter Bis Mittelschwerer Depressionen: Vergleich der Antidepressiven Wirksamkeit von Zwei Unterschiedlichen Dosierungsschemata," *Phytotherapie-Kongress*, 1993.

Bjerkenstedt, L., G. V. Edman, R. G. Alken, and M. Mannel, "Hypericum Extract LI 160 and Fluoxetine in Mild to Moderate Depression: A Randomized, Placebo-Controlled Multi-Center Study in Outpatients," *European Archives of Psychiatry and Clinical Neuroscience,* Vol. 255, No. 1, February 2005, pp. 40–47. PMID: 15538592

Bracher, A., "Johanniskraut 1x Täglich: Wirkung Klinisch Belegt/HYP611-Studie Bestätigt Wirksamkeit," *Verlagsbeilage Ärztliche Praxis*, Vol. 51, 2001, pp. 1–4.

Brenner, R., V. Azbel, S. Madhusoodanan, and M. Pawlowska, "Comparison of an Extract of Hypericum (LI 160) and Sertraline in the Treatment of Depression: A Double-Blind, Randomized Pilot Study," *Clinical Therapies,* Vol. 22, No. 4, Apr, 2000, pp. 411–419. PMID: 10823363

Egger, M., G. D. Smith, M. Schneider, and C. Minder, "Bias in Meta-Analysis Detected by a Simple, Graphical Test," *BMJ*, Vol. 315, 1997, pp. 629–634.

Fava, M., J. Alpert, A. A. Nierenberg, D. Mischoulon, M. W. Otto, J. Zajecka, H. Murck, and J. F. Rosenbaum, "A Double-Blind, Randomized Trial of St John's Wort, Fluoxetine, and Placebo in Major Depressive Disorder," *Journal of Clinical Psychopharmacology,* Vol. 25, No. 5, October 2005, pp. 441–447. PMID: 16160619

Gastpar, M., A. Singer, and K. Zeller, "Efficacy and Tolerability of Hypericum Extract STW3 in Long-Term Treatment with a Once-Daily Dosage in Comparison with Sertraline," *Pharmacopsychiatry,* Vol. 38, No. 2, March 2005, pp. 78–86. PMID: 15744631

———, "Comparative Efficacy and Safety of a Once-Daily Dosage of Hypericum Extract STW3-VI and Citalopram in Patients with Moderate Depression: A Double-Blind,

Randomised, Multicentre, Placebo-Controlled Study," *Pharmacopsychiatry,* Vol. 39, No. 2, March 2006, pp. 66–75. PMID: 16555167

Greenland, S., and M. P. Longnecker, "Methods for Trend Estimation from Summarized Dose-Response Data, with Applications to Meta-Analysis," *American Journal of Epidemiology,* Vol. 135, No. 11, 1992, pp. 1301–1309. PMID: 1626547

Hamling, J., P. Lee, R. Weitkunat, and M. Ambuhl, "Facilitating Meta-Analyses by Deriving Relative Effect and Precision Estimates for Alternative Comparisons from a Set of Estimates Presented by Exposure Level or Disease Category," *Statistics in Medicine,* Vol. 27, No. 7, March 30, 2008, pp. 954–970. PMID: 17676579

Hänsgen, K. D., J. Vesper, and M. Ploch, "Multicenter Double-Blind Study Examining the Antidepressant Effectiveness of the Hypericum Extract LI 160," *Journal of Geriatric Psychiatry and Neurology,* Vol. 7, Suppl. 1, 1994, pp. S15–S18. PMID: 7857501

Harrer, G., W. D. Hübner, and H. Podzuweit, "Effectiveness and Tolerance of the Hypericum Extract LI 160 Compared to Maprotiline: A Multicenter Double-Blind Study," *Journal of Geriatric Psychiatry and Neurology,* Vol. 7, Suppl. 1, 1994, pp. S24–S28. PMID: 7857503

Harrer, G., U. Schmidt, U. Kuhn, and A. Biller, "Comparison of Equivalence Between the St. John's Wort Extract LoHyp-57 and Fluoxetine," *Arzneimittelforschung,* Vol. 49, No. 4, April 1999, pp. 289–296. PMID: 10337446

Hartung, J., "An Alternative Method for Meta-Analysis," *Biometrical Journal,* Vol. 41, No. 8, 1999, pp. 901–916. PMID: 15206538

Hartung, J., and G. Knapp, "A Refined Method for the Meta-Analysis of Controlled Clinical Trials with Binary Outcome," *Statistics in Medicine,* Vol. 20, No. 24, December 30, 2001, pp. 3875–3889. PMID: 11782040

HDTSG—*See* Hypericum Depression Trial Study Group.

Higgins, J. P., D. G. Altman, P. C. Gotzsche, P. Juni, D. Moher, A. D. Oxman, J. Savovic, K. F. Schulz, L. Weeks, J. A. Sterne, Cochrane Bias Methods Group, and Cochrane Statistical Methods Group, "The Cochrane Collaboration's Tool For Assessing Risk of Bias in Randomised Trials," *BMJ,* Vol. 343, 2011, p. d5928. PMID: 22008217

Hypericum Depression Trial Study Group, "Effect of Hypericum Perforatum (St John's Wort) in Major Depressive Disorder: A Randomized Controlled Trial," *JAMA,* Vol. 287, No. 14, 2002, pp. 1807–1814. PMID: 11939866

Institute of Medicine, *Treatment for Posttraumatic Stress Disorder in Military and Veteran Populations: Initial Assessment.*, Washington, D.C., 2012.

IntHout, J., J. P. Ioannidis, and G. F. Borm, "The Hartung-Knapp-Sidik-Jonkman Method for Random Effects Meta-Analysis Is Straightforward and Considerably Outperforms the Standard DerSimonian-Laird Method," *BMC Medical Research Methodology,* Vol. 14, 2014, p. 25. PMID: 24548571

Kalb, R., R. D. Trautmann-Sponsel, and M. Kieser, "Efficacy and Tolerability of Hypericum Extract WS 5572 Versus Placebo in Mildly to moderately Depressed Patients. A Randomized Double-Blind Multicenter Clinical Trial," *Pharmacopsychiatry*, Vol. 34, No. 3, May 2001, pp. 96–103. PMID: 11434406

Kasper, S., I. G. Anghelescu, A. Szegedi, A. Dienel, and M. Kieser, "Superior Efficacy of St John's Wort Extract WS 5570 Compared to Placebo in Patients with Major Depression: A Randomized, Double-Blind, Placebo-Controlled, Multi-Center Trial [ISRCTN77277298]," *BMC Medicine*, Vol. 4, 2006, p. 14. PMID: 16796730

Kasper, S., H. P. Volz, H. J. Moller, A. Dienel, and M. Kieser, "Continuation and Long-Term Maintenance Treatment with Hypericum Extract WS 5570 After Recovery from an Acute Episode of Moderate Depression—A Double-Blind, Randomized, Placebo Controlled Long-Term Trial," *European Neuropsychopharmacology,* Vol. 18, No. 11, November 2008, pp. 803–813. PMID: 18694635

Knuppel, L., and K. Linde, "Adverse Effects of St. John's Wort: A Systematic Review," *Journal of Clinical Psychiatry,* Vol. 65, No. 11, 2004, pp. 1470–1479. PMID: 15554758

Laakmann, G., A. Dienel, and M. Kieser, "Clinical Significance of Hyperforin for the Efficacy of Hypericum Extracts on Depressive Disorders of Different Severities," *Phytomedicine,* Vol. 5, No. 6, December 1998, pp. 435–442. PMID: 23196026

Lecrubier, Y., G. Clerc, R. Didi, and M. Kieser, "Efficacy of St. John's Wort Extract WS 5570 in Major Depression: A Double-Blind, Placebo-Controlled Trial," *American Journal of Psychiatry,* Vol. 159, No. 8, August 2002, pp. 1361–1366. PMID: 12153829

Lenoir, S., F. H. Degenring, and R. Saller, "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," *Phytomedicine,* Vol. 6, No. 3, 1999, pp. 141–146. PMID: 10439477

Lewin Group and ECRI Institute, *Management of Dyslipidemia: Evidence Synthesis Report. Clinical Practice Guideline*, Washington, D.C.: Veterans Health Administration, U.S. Department of Veterans Affairs and the U.S. Department of Defense, 2014.

Linde, K., M. M. Berner, and L. Kriston, "St John's Wort for Major Depression," *Cochrane Database of Systematic Reviews,* Vol. CD000448, No. 4, 2008. PMID: 18843608

Linde, K., L. Kriston, G. Rücker, S. Jamil, I. Schumann, K. Meissner, K. Sigterman, and A. Schneider, "Efficacy and Acceptability of Pharmacological Treatments for Depressive Disorders in Primary Care: Systematic Review and Network Meta-Analysis," *Annals of Family Medicine*, Vol. 13, No. 1, January–February 2015, pp. 69–79.

Liu, F. F., C. Y. W. Ang, T. M. Heinze, J. D. Rankin, R. D. Beger, J. P. Freeman, and J. O. Lay, Jr., "Evaluation of Major Active Components in St. John's Wort Dietary Supplements by High-Performance Liquid Chromatography with Photodiode Array Detection and Electrospray Mass Spectrometric Confirmation," *Journal of Chromatography A*, 2000, pp. 85–92. PMID: 10949475

Liu, J., Z. W. Meng, L. Y. Li, L. S. Feng, and H. Yang, "Effect of St. John's Wort Extract on Depressive Disorder in Elderly Patients with Unstable Angina," *World Journal of Emerging Medicine,* Vol. 1, No. 1, 2010, pp. 41–44. PMID: 25214939

Mannel, M., U. Kuhn, U. Schmidt, M. Ploch, and H. Murck, "St John's Wort Extract LI160 for the Treatment of Depression with atypical features—A Double-Blind, Randomized, and Placebo-Controlled Trial," *Journal of Psychiatric Research,* Vol. 44, No. 12, 2010, pp. 760–767. PMID: 20181361

McPherson, F., and M. A. Schwenka, "Use of Complementary and Alternative Therapies Among Active Duty Soldiers, Military Retirees, and Family Members at a Military Hospital," *Military Medicine,* Vol. 169, No. 5, 2004, pp. 354–357. PMID: 15185998

Montgomery, S. A., W. D. Hübner, and H. G. Grigoleit, "Efficacy and Tolerability of St. John's Wort Extract Compared with Placebo in Patients with a Mild to Moderate Depressive Disorder," *Phytomedicine,* Vol. 7, Suppl. 2, 2000, p. 107.

Moreno, R. A., C. T. Teng, K. M. de Almeida, and H. Tavares Jr., "Hypericum Perforatum Versus Fluoxetine in the Treatment of Mild to Moderate Depression: A Randomized Double-Blind Trial in a Brazilian Sample," *Revista Brasileira de Psiquiatria,* Vol. 28, No. 1, 2006, pp. 29–32. PMID: 16612487

Orsini, N., R. Li, A. Wolk, P. Khudyakov, and D. Spiegelman, "Meta-Analysis for Linear and Nonlinear Dose-Response Relations: Examples, an Evaluation of Approximations, and Software," *American Journal of Epidemiology,* Vol. 175, No. 1, 2012. PMID: 22135359

Pakseresht, S., H. Boustani, M. E. Azemi, J. Nilsaz, R. Babapour, and M. R. Haghdust, "Evaluation of Pharmaceutical Products of St. John's Wort Efficacy Added on Tricyclic Antidepressants in Treating Major Depressive Disorder: A Double Blind Randomized Control Trial," *Jundishapur Journal of National Pharmaceutical Products,* Vol. 7, No. 3, Summer 2012, pp. 106–110. PMID: 24624165

Philipp, M., R. Kohnen, and K. O. Hiller, "Hypericum Extract Versus Imipramine or Placebo in Patients with Moderate Depression: Randomised Multicentre Study of Treatment for Eight Weeks," *BMJ,* Vol. 319, No. 7224, December 11, 1999, pp. 1534–1538. PMID: 10591711

Rahman, R., M. A. Ansari, Z. Hayder, A. A. Siddiqui, I. A. Bukhari, and M. A. Qayyum, "Double Blind Placebo Controlled Clinical Trial Examining the Effectiveness of St. John's Wort (Hypericum Perforatum) in Mild to Moderate Depression," *Journal of Pakistan Psychiatric Society,* Vol. 5, No. 2, 2008, pp. 105–111.

Sánchez-Meca, J. , and F. Marín-Martínez, "Confidence Intervals for the Overall Effect Size in Random-Effects Meta-Analysis," *Psychological Methods,* Vol. 13, No. 1, March 2008, pp. 31–48. PMID: 18331152

Schrader, E., "Equivalence of St John's Wort Extract (Ze 117) and Fluoxetine: A Randomized, Controlled Study in Mild-Moderate Depression," *International Clinical Psychopharmacology,* Vol. 15, No. 2, March 2000, pp. 61–68. PMID: 10759336

Schrader, E., B. Meier, and A. Brattström, "Hypericum Treatment of Mild–Moderate Depression in a Placebo–Controlled Study: A Prospective, Double–Blind, Randomized, Placebo–Controlled, Multicentre Study," *Human Psychopharmacology: Clinical & Experimental,* Vol. 13, No. 3, 1998, pp. 163–169. PMID: 11822078

Shelton, R. C., M. B. Keller, A. Gelenberg, D. L. Dunner, R. Hirschfeld, M. E. Thase, J. Russell, R. B. Lydiard, P. Crits-Cristoph, R. Gallop, L. Todd, D. Hellerstein, P. Goodnick, G. Keitner, S. M. Stahl, and U. Halbreich, "Effectiveness of St John's Wort in Major Depression: A Randomized Controlled Trial," *JAMA,* Vol. 285, No. 15, April 18, 2001, pp. 1978–1986. PMID: 11308434

Sidik, K., and J. N. Jonkman, "Robust Variance Estimation for Random Effects Meta-Analysis," *Computational Statistics & Data Analysis,* Vol. 50, No. 12, 2006, pp. 3681–3701.

Szegedi, A., R. Kohnen, A. Dienel, and M. Kieser, "Acute Treatment of Moderate to Severe Depression with Hypericum Extract WS 5570 (St John's Wort): Randomised Controlled Double Blind Non-Inferiority Trial Versus Paroxetine," *BMJ,* Vol. 330, No. 7490, March 5, 2005, p. 503. PMID: 15708844

Uebelhack, R., J. Gruenwald, H. J. Graubaum, and R. Busch, "Efficacy and Tolerability of Hypericum Extract STW 3-VI in Patients with Moderate Depression: A Double-Blind, Randomized, Placebo-Controlled Clinical Trial," *Advances in Therapy,* Vol. 21, No. 4, July–August 2004, pp. 265–275. PMID: 15605620

U.S. Department of Veterans Affairs and U.S. Department of Defense, *VA/DoD Clinical Practice Guideline for Management of Major Depressive Disorder (MDD)*, Washington, D.C.: Management of MDD Working Group, 2009.

U.S. Preventive Services Task Force, *US Preventive Services Task Force Procedure Manual*, Rockville, Md., 2008.

Ustun, T. B., J. L. Ayuso-Mateos, S. Chatterji, C. Mathers, and C. J. L. Murray, "Global Burden of Depressive Disorders in the Year 2000," *British Journal of Psychiatry,* Vol. 184, 2004, pp. 386–392. PMID: 15123501

van Gurp, G., G. B. Meterissian, L. N. Haiek, J. McCusker, and F. Bellavance, "St John's Wort or Sertraline? Randomized Controlled Trial in Primary Care," *Canadian Family Physician,* Vol. 48, 2002, pp. 905–912. PMID: 12053635

Volz, H. P. , R. Eberhardt, and G. Grill, "Efficacy and Tolerance of the St John's Wort Extract D-0496 in Mild to Moderate Depression: A Placebo-Controlled, Double-Blind 6-Week Trial [Wirksamkeit und Verträglichkeit des Johanniskrautextraktes D–0496 bei Leichten bis Mittelschweren Depressiven Episoden: Plazebokontrollierte Doppelblindstudie über 6 Wochen]," *Nervenheilkunde,* Vol. 19, 2000, pp. 401–405.

Vorbach, E. U., K. H. Arnoldt, and W. D. Hubner, "Efficacy and Tolerability of St. John's Wort Extract LI 160 Versus Imipramine in Patients with Severe Depressive Episodes According to ICD-10," *Pharmacopsychiatry,* Vol. 30, Suppl. 2, September 1997, pp. 81–85. PMID: 9342765

Wheatley, D., "LI 160, an Extract of St. John's Wort, Versus Amitriptyline in Mildly to Moderately Depressed Outpatients—A Controlled 6-Week Clinical Trial," *Pharmacopsychiatry,* Vol. 30, Suppl. 2, September 1997, pp. 77–80. PMID: 9342764

Witte, B., G. Harrer, T. Kaptan, H. Podzuweit, and U. Schmidt, "[Treatment of Depressive Symptoms with a High Concentration Hypericum Preparation: A Multicenter Placebo-Controlled Double-Blind Study]," *Fortschritte der Medizin,* Vol. 113, No. 28, October 10, 1995, pp. 404–408. PMID: 7498868

Woelk, H., "Comparison of St John's Wort and Imipramine for treating Depression: Randomised Controlled Trial," *BMJ,* Vol. 321, No. 7260, September 2, 2000, pp. 536–539. PMID: 10968813

World Health Organization, *Depression (Fact Sheet N. 369)*, 2012. As of August 12, 2015: http://www.who.int/mediacentre/factsheets/fs369/en/