Methods

This is a systematic literature review and meta-analysis. The protocol has been published on Prospero (Prospero Record Registration No.: CRD42013004407).

Literature Search

We employed the Cochrane Highly Sensitive Search Strategy(22) and searched MEDLINE (via PubMed), PsycINFO, EMBASE and the Cochrane Central Register of Controlled Trials (CENTRAL) through march 2014 without language and time restrictions. We applied the sensitivity- and precision-maximizing version and the search terms used by CENTRAL to identify randomized controlled trials in EMBASE as described in detail in the Cochrane Collaboration Handbook(22), adapted to controlled trials. Together with trial filters, we used generic terms for depressive symptoms as well as specific diagnoses of affective disorders, combined with generic terms for combined treatment and individual drug names (search entry available online: see Annex 1). In addition, reference lists of included trials were hand-searched.

Eligibility criteria

Trials were included when they met the following criteria: the existence of a control group of antidepressant monotherapy(including open-label and non-randomized trials), inclusion of participants aged 18 or older, of both sexes with depressive disorder, diagnosis according to standard operationalized criteria, such as Feighner Criteria, Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV, DSM-IV-TR, CCMD, ICD-10, or similar approaches detailed in the paper. Concurrent diagnoses of other psychiatric disorders as well as co-morbid

medical conditions were no exclusion criteria. Studies specifically on bipolar depression were excluded. Only trials assessing severity of depression with a standardized and established rating scale (e.g. HAMD, MADRS, CGI, BDI) were considered relevant. Pharmacological intervention was to be followed over a time period of at least two weeks (for the current episode of depression) prior to final assessment.

Irrespective of dosage, we included all pharmacological interventions using a combination of two antidepressants, i.e. initial combination therapy as well as adjunctive administration of a second to a first antidepressant. Both trials on first-line treatment and trials among patients with resistance to previous antidepressive treatment(s) were included. We excluded trials on maintenance therapy.

Data collection

Titles and abstracts of all studies retrieved by the search process were reviewed for inclusion. When necessary, full texts were obtained in order to assess eligibility. In unclear cases, inclusion or exclusion was decided in a discussion among all authors (JH, TB, CB). Included full texts were read independently by two reviewers (JH, TB, CB). Using Cochrane's risk of bias tool(22) as a guideline, two reviewers independently assessed risk of bias of each included trial. Trials were judged as holding „high risk", „low risk" or „unclear risk" of bias, taking into account the following specific domains: random sequence generation, allocation concealment, blinding of participants, personnel, and outcome assessors, incomplete outcome data, selective reporting, sponsorship and other potential sources of bias. An overall assessment of risk of bias was added in order to constitute subgroups of different methodological rigor. Disagreements were solved via discussion with the third reviewer.
All necessary data to evaluate outcomes were retrieved independently by two reviewers, using a standardized data extraction form in accordance with the Cochrane Collaboration

Handbook.(22) If full texts were not available, the authors of the articles were contacted for further information.

Outcome criteria

Primary outcome

The primary outcome criterion was standardized mean difference (SMD) between combination and monotherapy groups. As efficacy assessment of antidepressant treatment varies among studies, we combined different assessment instruments by calculating SMD and SE (standard error) from means or Odds Ratios (OR), depending on the parameters of outcome reported (e.g. remission, response, mean differences of rating scale scores at study endpoint). In case of more than one outcome measurement, parameters were selected according to the following hierarchy:

1. Score on depression rating scale at study endpoint (+/- standard deviation (SD))
2. Remission: defined as scores below pre-determined thresholds on a depression scale (e.g. </= 8 on the 17-item HDRS or the MADRS). We adopted trial author's definition.
3. Response: defined as a decrease on a depression rating scale (e.g. at least 50 % on the 17-item HDRS or the MADRS). We adopted trial author's definition.

Secondary outcomes

Pre-specified secondary outcomes, were: Remission rates, response rates, difference in depression rating at study endpoint (SMD), and tolerability documented as dropouts due to any reason and dropouts due to adverse effects (AE).

Subgroup analyses

Pre-specified subgroup analyses included samples of studies that were randomized and double-blind, had a low risk of bias, and included non-responders only.

Additional moderator analyses

In addition to effect size calculations for subgroups moderator analyses aimed at elucidating the role of possible confounders. Specifically, in random effects meta regression associations were analyzed of SMD with duration of follow up and with both the difference between and ratio of imipramine equivalent doses (as provided in(23), and, in one case, by personal communication with Ross J. Baldessarini in January 2014) in combination and monotherapy arms. Further analyses included comparisons of dichotomized possible confounders: randomized versus non-randomized treatment allocation, low versus non-low risk of bias studies, and treatment resistant versus non-treatment resistant status.

Post-hoc analyses

In post-hoc analyses we evaluated different treatment modalities (comparison of combination versus either continuation of monotherapy or switch to another monotherapy), analyzed studies using a certain combination of antidepressants (combination of a monoamine reuptake inhibitor (SSRI, SNRI, TCA) with antagonists of pre-synaptic α2-auto-receptors (mianserin, mirtazapine, trazodone) versus other combinations), or restricted the analysis to studies excluding patients with bipolar disorder or to studies of MDD patients only.

Data analysis

Data analyses were carried out using the Review Manager (RevMan 5.2.1) as well as Comprehensive Meta-Analysis (Version 2) software and according to the Cochrane Collaboration Handbook.(22) We used ITT-populations to assess outcomes, taking into account the total number of randomized patients or, in non-randomized trials, all patients initially included in each group. Even in the absence of statistically significant heterogeneity SMDs and odds ratios (with 95% confidence intervals) were calculated with a random-effects

model because the studies in this analysis differed in several methodological aspects, such as randomization, blinding, or diagnostic criteria and measurement scales used. Heterogeneity among studies was assessed by I^2 – and additional Tau^2–statistics, as the former is known to become inflated with increasing sample size(24).

If studies consisted of more than one comparison (i.e. either more than one monotherapy group, or more than one combination group), following recommendations of the Cochrane Collaboration Handbook,(22) we combined intervention groups to avoid ‚double-counts' of patients: Outcome data of multiple groups of one intervention (combination or monotherapy) were pooled and correspondent SDs were calculated.

The likelihood of publication bias with regard to the primary outcome was assessed by a funnel plot of studies. Power calculations were conducted with G*Power 3.(25) In addition, Egger's test, a trim and fill procedure and a fail-safe N calculation (Orwin's) were carried out.

The robustness of the primary outcome was tested by removing all studies one by one from the analysis.

For the primary outcome, statistical significance was set at an alpha of 0.05. For all secondary outcomes and for all analyses of subgroups p-values are presented, but not as a marker of statistical significance.