Systematic Review or Meta-Analysis

In search of a dose–response relationship in SSRIs—a systematic review, meta-analysis, and network meta-analysis

Braun C, Adams A, Rink L, Bschor T, Kuhr K, Baethge C. In search of a dose–response relationship in SSRIs—a systematic review, meta-analysis, and network meta-analysis.

Objective: Recent meta-analyses on dose–response relationships of SSRIs are largely based on indirect evidence. We analyzed RCTs directly comparing different SSRI doses.

Method: Systematic literature search for RCTs. Two raters independently screened articles and extracted data. Across SSRIs, doses defined as low, medium, and high doses, based on drug manufacturers’ product monographs, were analyzed in pairwise random-effects meta-analyses and in a sensitivity network meta-analysis with regard to differences in antidepressive efficacy (primary outcome). We also analyzed all direct comparisons of different dosages of specific SSRIs. (Prospero CRD42018081031).

Results: Out of 5333 articles screened, we included 33. Comparisons of dosage groups (low, medium, and high) resulted in only small and clinically non-significant differences for SSRIs as a group, the strongest relating to medium vs low doses (SMD: −0.15 [95%-CI: −0.28; −0.01]) and not sustained in a sensitivity analysis. Among different doses of specific SSRIs, no statistically significant trend emerged for efficacy at higher doses, but 60 mg/day fluoxetine are statistically significantly inferior to 20 mg/day. Paroxetine results are inconclusive: 10 mg/day are inferior to higher doses, but 30 and 40 mg/day are inferior to 20 mg/day. Meaningful effects cannot be ruled out for certain drugs and dosages, often investigated in only one trial. Dropout rates increase with dose—particularly due to side effects. Network meta-analyses supported our findings.

Conclusions: There is no conclusive level I or level II evidence of a clinically meaningful dose–response relationship of SSRIs as a group or of single substances. High SSRI doses are not recommended as routine treatment.

Summations

- We found no positive and consistent evidence of a statistically or clinically significant dose–response relationship in SSRIs as a group.
- By a small margin (SMD of 0.2), fluoxetine 60 mg/day is statistically and clinically significantly inferior to 20 mg.
- Higher doses of SSRI come with more dropouts, overall, and due to adverse events.

Limitations

- Some dose comparisons of specific SSRIs have not been comprehensively studied.
- For several comparisons, confidence intervals remain wide and heterogeneity was substantial.
- Many studies carry a high or unclear risk of bias.
Introduction

Selective serotonin reuptake inhibitors (SSRIs) constitute a large part of antidepressants employed in the treatment of depressive disorders. In Germany, for example, SSRI prescriptions increased about 80% between 2007 and 2017 (1).

In our clinical experience, SSRI doses often are increased above minimal dosages recommended—either from the beginning of treatment or in its course. Still, it is uncertain whether there is a dose–response relationship for SSRIs as a group or for specific SSRIs in antidepressive treatment. In randomized controlled trials (RCTs), Montgomery and co-authors (2), for example, reported such a relationship, while others, for example, Burke et al. (3) and Fabre et al. (4), did not.

Recently, the debate has been sparked by meta-analyses with differing conclusions. Jakubovski et al. (5) concluded that higher doses of SSRIs were superior to lower doses. In a network meta-analysis of SSRIs and other second-generation antidepressants, Furukawa et al. (6) came to the opposite conclusion that higher doses are not superior and advised against their use. Both meta-analyses, however, employed indirect comparisons, be it by analyzing different SSRI doses across studies, not within studies (5), or by choosing the response against placebo as a yardstick (6). While analyzing indirect comparisons, for example, via placebo, certainly is informative and allows the inclusion of a great many studies, it is fraught with even more problems regarding comparability of studies than pairwise meta-analysis. In order to minimize this source of variance we favor direct comparisons of different dosages within randomized trials and employ network meta-analysis only as a sensitivity analysis providing a plausibility check. To our knowledge, the most recent meta-analyses comprehensively summarizing direct randomized comparisons of SSRI dosages date back to the turn of the century (7,8). In the meantime, several new studies on the subject have been published, for example, Jefferson et al. (9); Trivedi et al. (10); Yevtushenko et al. (11), Rapaport et al. (12); Hirayasu et al. (13,14), Mathews et al. (15) and Kato (16).

Consequently, we conducted a systematic literature search of RCTs comparing two or more doses of an SSRI and summarized the findings by pairwise meta-analyses.

Aims of the study

The primary objective of our analyses is to determine whether there is a dose–response relationship in antidepressant treatment with SSRIs as a group or for specific SSRIs. We also examined the association between dose and dropout-rate.

Methods

This is a registered systematic literature review and meta-analysis (PROSPERO no. CRD42018081031) reported according to the PRISMA guidelines. This analysis is part of a larger study on dose–response relationships in antidepressants. The present manuscript is restricted to SSRIs. We follow the drug classification of the FDA (17).

Literature search and data extraction

We conducted literature search and data extraction according to the Cochrane Handbook for Systematic Reviews of Interventions (18). Search terms and their combinations in Boolean search are detailed in the supplement (Table S1). Broadly, we used generic and specific terms to find articles on randomized controlled trials, depressive disorders, and dose–response relationships of antidepressants. No restrictions applied to language, publication date and status, gender, age, ethnicity, or gray literature. We searched Embase database until February 2017 and PsycINFO, MEDLINE, and Cochrane’s Central databases until June 2019. We screened reference lists of 125 reviews, meta-analyses, pooled analyses, and all articles eventually included. Two authors (CBr and LR) independently screened all titles and abstracts, and independently read full texts of all studies considered potentially eligible. Disagreements were solved by discussion (CBa, CBr, and LR).

Inclusion criteria. We included articles meeting the following criteria: (i) RCTs with randomization to at least two different doses of one SSRI (citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline, or vilazodone). (ii) Patients diagnosed with depressive disorders according to established diagnostic systems, such as DSM or ICD. (iii) Treatment duration of at least three weeks. (iv) Outcome assessment by an established instrument, for example, HAMD, MADRS, and BDI.

We excluded studies investigating different doses of antidepressants prescribed consecutively or at different stages during treatment, as in dose increase studies of second-step strategies. We also excluded studies that, after non-response, compared continuation of a dose with a dose increase arm (19), because, in such trials, participants are exposed to doses for different periods of time (e.g., Licht, Qvitza (20)). Finally, we excluded studies on secondary depressive syndromes, for example, in Parkinson’s disease.
Data extraction. Two authors (CBr, LR) independently extracted data using a pre-specified, standardized, Excel-based data extraction form following the Cochrane Collaboration Handbook (18) and similar to the forms used in earlier studies by our group (19,21). We contacted authors by e-mail if relevant data were missing. If data were available from graphs only, two authors independently (CBr, LR) used Plot digitizer (SourceForge Project, sourceforge.net) to read out numbers.

Risk of bias. For risk of bias assessment, we used the Cochrane Collaboration Handbook risk of bias tool (chap. 8.5) (18): Two authors (CBr, LR) rated all studies regarding random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other sources of bias. Investigations must show ‘low’ risk of bias in at least five out of six mentioned items to be considered a low risk of bias study. Otherwise, they were classified as ‘high risk’ or as having an ‘unclear risk’.

Data analysis

Primary outcome. Primary outcome was the comparison of efficacy between different doses as expressed in differences of psychopathology scores. For studies reporting more than one depression rating scale, we used hierarchically 1. HDRS, 2. MADRS, 3. BMRS, 4. other scales. As effect variable, we prioritized change from baseline over follow-up value and other endpoints.

If authors did not provide depression scores, we calculated SMDs from other outcomes, such as 1. response rate or, 2. remission rate, defined as a decrease on a depression rating scale (following triallists’ definition). For the purposes of this analysis, we defined a standardized mean difference (SMD) of 0.2—corresponding to a small effect size in the framework of Cohen (22)—as the threshold to a clinically meaningful effect.

We preferred intention-to-treat (ITT) data and followed the authors’ method of replacing missing data, for example, LOCF or mixed-effect model with repeated measure (MMRM).

Secondary outcome. We analyzed dropouts due to side effects and due to any reason. We decided a priori not to analyze dropouts due to inefficacy because in our impression this parameter is often not reported in RCTs. Also, in order to keep analyses to a reasonable number and to avoid problems of multiplicity we decided against analyzing responder and remitter rates separately.

Subgroup and sensitivity analysis, and posthoc analyses. As a sensitivity analysis, placebo arms were added to active treatment arms in network meta-analysis. We also added or deleted studies from the analysis to test its robustness if studies appeared to contribute particularly to high heterogeneity (inconsistency).

Publication bias. We examined publication bias (primary outcome): In pairwise comparisons with at least ten trials, we examined the presence of small-study effects graphically by funnel plots and used linear regression tests (23) to test funnel plot asymmetry (p threshold 0.1) (24).

Data synthesis. As efficacy measurement, we used standardized mean differences and 95% confidence intervals (CI). If original scores were available, we used change from baseline, otherwise, we calculated them from baseline and final scores (chap. 7.7 on extracting and converting study results) (18). Alternatively, we converted dichotomous measures into SMD (chap. 9.4 on summarizing effects) (18). Missing standard deviations (SD) were calculated from CI or P-values (chap. 7.7) (18), or imputed using SDs from baseline and final scores assuming a correlation coefficient ($r$) of 0.5 (chap 16.1 on missing data) or imputed from similar studies according to a validated imputation method (25). We used only studies reporting SD of mean change to calculate pooled SDs.

Pairwise meta-analyses. We followed two approaches to investigate dose–response relationships. In one approach, we graded SSRIs dosages as low, medium, and high according to the manufacturers’ drug information (product monographs, references available upon request). We defined the highest dose recommended by the product monographs as high, as well as all doses above. All other doses recommended by the product monographs for clinical use we defined as medium, whereas all doses below we defined as low. For example, the drug manufacturer of sertraline recommended daily dosages between 50 and $\lessapprox$ 200 mg (all dosages are mg per day) for clinical use (26). Therefore, we defined the upper dose (300 mg) tested in the study by Guy et al. in 1986 (27) as high. In the same vein, we considered 12.5 mg per day of paroxetine in Rapaport et al. study (2009) (12) a low dose, and all doses of fluoxetine between 20 and 79 mg as medium doses (Table S2). Posthoc, in light of the studies included, we decided for
a sensitivity analysis without studies with very large effects as well as those for which we had to impute effects using the Bucher method (28).

In another approach, we analyzed all dose comparisons. While 20, 40, and 60 mg of fluoxetine are all considered medium doses as per our first approach, it is instructive to compare those doses because they represent common dosages in routine treatment. We listed all studies comparing specific doses, but restricted meta-analyses to comparisons in at least two studies.

In all pairwise comparisons, including dropout analyses, we applied random-effects meta-analyses (DerSimonian & Laird) due to design differences among studies. Heterogeneity is expressed by the $I^2$ statistic and reported throughout. In our main analysis, we did carry out a sensitivity analysis leaving out highly inconsistent studies if enough studies were available. Analyses were carried out with Review Manager and Comprehensive Meta-Analysis 2. Since we calculated many summary estimates the risk of type I errors is inflated, and we consider only the two main analyses of our two approaches as confirmatory, all other p-values are exploratory.

**Network meta-analysis (NMA).** As a sensitivity analysis to the first approach, we carried out a network meta-analysis comparing low, medium, and high doses of SSRIs. In addition, we carried out a NMA with placebo arms from direct comparison studies added to the network. Whenever data were sufficiently similar to be combined, we performed a NMA using the frequentist weighted least squared approach (29). We used a random-effects model, taking into account the correlated treatment effects in multi-arm studies and assumed a common estimate for the heterogeneity variance across different comparisons. We created network plots using the CINeMA software (30) and evaluated heterogeneity and inconsistency using the generalized heterogeneity statistic $Q_{\text{total}}$ and the generalized $I^2$ statistic (31). We used the `decomp.design` command in the R package netmeta (32,33) . To evaluate the presence of inconsistency locally, we compared direct and indirect treatment estimates of each treatment comparison (34) using the `netsplit` command in the R package netmeta (32,33).

**PICO Statement.** This is a systematic review, pairwise meta-analysis, and network meta-analysis to investigate whether, in RCTs (design), different doses of SSRIs (intervention, control) show different efficacy, as measured by standardized mean differences of psychopathology scores (outcome), in treating patients with depressive disorders (participants).

**Analysis carried after submission of the manuscript.** In order to carry out additional sensitivity analyses, we contrasted (i) minimum recommended SSRI doses to all SSRI doses above and (ii) all SSRI doses below minimum recommended doses to all SSRI doses above.

**Results**

Our literature search netted 6439 articles. Elimination of duplicates resulted in 5333 articles. After screening titles and abstracts, we evaluated 205 papers at full-text level, and eventually included 33 papers for quantitative analysis (Fig. 1). Seven articles focused on fluoxetine, seven on citalopram, six each on sertraline and paroxetine, four on escitalopram, three on vilazodone, and one on fluvoxamine (not exclusive).

Thirty-three doses of specific SSRIs were compared 56 times. Twenty-nine out of 33 trials were double-blind. In comparison to the papers by Bollini et al. (7) and Baker et al. (8), the most recent meta-analyses of RCTs directly comparing different doses of SSRIs, our analysis is based on 26 additional studies. Twenty-two studies were not included in either of those two meta-analyses (7,8). In total, the studies selected included 12 052 patients, 9454 on various SSRI doses, and 2598 on placebo. Based on the smaller samples of studies presenting data on gender ($n = 25$) and age ($n = 26$), 58% of trial participants were women and the average age was 41.8 years (weighted grand mean). Except for some older studies, patients were included in the trials based on a diagnosis of major depression (27/33 studies), and most studies excluded bipolar patients (23/31 studies presenting information).

**Pairwise meta-analysis of low, medium, and high doses**

In 22 studies, we found 26 pairwise comparisons of low, medium, and high doses (Table 1; Fig. 2): medium vs. low dose studies resulted in a small, but statistically significant standardized mean difference of $-0.15$ [-0.28; -0.01] in favor of medium doses, with considerable heterogeneity ($I^2$: 63%). High doses were numerically slightly superior to low doses, but not statistically significantly (SMD $-0.06$ [-0.22; 0.09], $I^2$: 0%). The comparison of medium and high doses revealed no difference (SMD $-0.00$ [-0.12; 0.12], $I^2$: 37%) (Table 1, Fig. 2). A funnel
plot (Fig. S1) and Egger’s test gave no indication of serious publication bias for the comparison of low vs medium doses, the only comparison with enough trials for meaningful publication bias analysis.

A sensitivity analysis, excluding studies by Guy (27), Ghose (35), Yevtushenko (11), SER101 (36), and SER310 (37) due to high inconsistency resulted in minor effects throughout, particularly a numerically smaller effect in the comparison of medium vs. low (SMD $-0.09 \ [-0.20; \ 0.02]$ $I^2$: 42%) (table 1).

**Dropouts.** Dropouts were more frequent in higher relative to lower doses, in particular those due to side effects. In medium vs low dose arms OR for dropouts due to adverse events was 1.70 ([1.29; 2.31]).

---

| Records identified through database searching (n = 6375) | Additional records identified through other sources (n = 64) |
|----------------------------------------------------------|----------------------------------------------------------|
| Records after duplicates removed (n = 5269)               |                                                          |
| Records screened (n = 5333)                               | Records excluded (n = 5128)                              |
| Full-text articles assessed for eligibility (n = 205)     | Full-text articles excluded, with reasons (n = 172)      |
|                                                          |   n= 62 other than SSRI                                   |
|                                                          |   n= 61 no randomised comparison of ≥ 2 different doses of one AD |
|                                                          |   n= 18 AD not licensed                                  |
|                                                          |   n= 10 missing data                                     |
|                                                          |   n= 9 patients reported in another selected paper       |
|                                                          |   n= 8 no acute treatment                                |
|                                                          |   n= 4 no continuous fixed dose                          |

Studies included in quantitative synthesis (meta-analysis) (n = 33)

---

*Fig. 1. PRISMA flow diagram [Colour figure can be viewed at wileyonlinelibrary.com]*
2.23] 10 studies; $I^2$: 5%) and in high-dose vs medium-dose arms 1.65 ([1.07; 2.54] 7 studies; $I^2$: 41%). Only one study reported OR on high versus low dose (38): OR 2.47 [1.20; 5.07] (Table 2).

Pairwise meta-analyses of all doses investigated

While all comparisons of the previous meta-analysis pertain to pre-specified dose ranges of low, medium, and high (e.g., fluoxetine < 20 mg vs. 20–79 mg vs. ≥ 80 mg), the following comparisons regard the analysis of all doses applied in clinical trials (e.g., fluoxetine 20 mg versus 40 mg vs. 60 mg). For ten comparisons of different dosages, we found at least two studies and therefore calculated a meta-analysis (Table 3). The total number of studies in this analysis was 27.

With little heterogeneity, paroxetine 20 mg seemed to be superior to 10 mg (SMD −0.24 [−0.37; −0.12], Table 3). Other comparisons of paroxetine doses were inconclusive, with some, particularly smaller, studies indicating a positive dose–response relationship—a finding not supported by all larger studies (Table 3). SMD for the comparison of fluoxetine 60 mg vs. fluoxetine

![Table 1. Pairwise analysis of primary outcome and sensitivity analysis (SSRI doses categorized as low, medium, high)](image)

| SMD† | 95%-CI | Comparisons (n) | Heterogeneity $I^2$ |
|------|--------|-----------------|---------------------|
| Medium vs. Low | −0.15 | −0.28; −0.01 | 12 | 63% |
| High vs. Low | −0.06 | −0.22; 0.09 | 3 | 0% |
| High vs. Medium | −0.00 | −0.12; 0.12 | 11 | 37% |

Sensitivity analysis†

| SMD† | 95%-CI | Comparisons (n) | Heterogeneity $I^2$ |
|------|--------|-----------------|---------------------|
| Medium vs. Low | −0.09 | −0.20; 0.02 | 10 | 42% |
| High vs. Low | −0.06 | −0.22; 0.09 | 3 | 0% |
| High vs. Medium | −0.03 | −0.15; 0.09 | 8 | 36% |

†A negative standardized mean difference (SMD) indicates a better effect of the higher dose.

‡Exclusion of studies Guy (27), Ghose (35), Yevtushenko (11), SER101 (36), and SER310 (37).

![Fig. 2. Meta-analysis of studies comparing SSRI doses categorized as low, medium, and high](image)
20 mg was 0.20 [0.01; 0.39] (no heterogeneity), showing the lower dose is more effective. Most comparisons (N = 7) were conducted for fluoxetine 40 mg versus fluoxetine 20 mg, resulting in a clinically and statistically non-significant SMD of -0.06 [-0.25; 0.12] in favor of the higher dose. Similarly, estimates of the other comparisons were not indicative of clinically relevant differences (Table 3).

Owing to a lack of studies, only few comparisons of specific doses suggested itself for meta-analyses regarding dropouts. As an exception, 20 mg of fluoxetine consistently caused fewer dropouts for any reason or dropouts due to side effects than 60 mg (OR for the latter 0.25 [0.14; 0.44] 3 studies, I²: 0%). In the same way, patients experienced fewer treatment terminating side effects under 10 mg paroxetine compared to 20 mg: OR 0.54 [0.28; 1.04] 4 studies, I²: 29%.

### Risk of bias analysis

The risk of bias analysis revealed high risk of bias in most studies (Table S5). Low risk of bias studies showed no discernible dose–response effects in citalopram (38) and vilazodone (15) treatment, and superiority of 20 mg paroxetine over 10 mg (10,12). Technically, Ghose’s study is a low risk of bias study (35) and showed superiority of 30 over 15 mg paroxetine, however, with only six patients

---

**Table 2. Dropouts, overall, and due to adverse events**

| Substances | Dose (mg/day) | n studies | n patients | Dropouts overall | OR (95% CI) | Comparisons (n) | Heterogeneity I² |
|------------|--------------|-----------|------------|-------------------|-------------|------------------|------------------|
| Fluvoxamine | 150 vs. 100 | 1 (63) | 121 vs. 119 | -0.04 [-0.29; 0.21] | 0.762 Unclear | 10 (63) | 150 vs. 100 |
| Fluvoxamine | 100 vs. 50 | 1 (4) | 1074 vs. 1144 | 0.08 [-0.17; -0.00] | 0.963 Unclear | 10 (4) | 100 vs. 50 |
| Fluvoxamine | 50 vs. 25 | 1 (61) | 8 vs. 8 | -0.25 [-1.23; 0.74] | 0.621 Unclear | 10 (61) | 50 vs. 25 |
| Citalopram | 60 vs. 20 | 1 (38) | 286 vs. 207 | 0.07 [0.55; 0.32] | 0.196 Unclear | 10 (38) | 60 vs. 20 |
| Citalopram | 50 vs. 25 | 1 (61) | 8 vs. 10 | -0.39 [-0.58; 0.18] | 0.898 Unclear | 10 (61) | 50 vs. 25 |
| Citalopram | 40 vs. 20 | 1 (38) | 8 vs. 10 | 0.91 [0.60; 0.22] | 0.963 Unclear | 10 (38) | 40 vs. 20 |
| Citalopram | 30 vs. 10 | 1 (63) | 8 vs. 10 | 0.551 High | 0.762 Unclear | 10 (63) | 30 vs. 10 |
| Citalopram | 20 vs. 10 | 1 (38) | 198 vs. 207 | 0.33 [0.26; 0.41] | 0.751 High | 10 (38) | 20 vs. 10 |

**Table 3. All efficacy comparisons of distinct SSRI dosages (in mg/day). If at least two comparisons were available, meta-analysis results are presented**

| Substance | Dose (mg/day) | n studies | n patients | Std diff in means [95% CI] | OR (95% CI) | Comparisons (n) | Heterogeneity I² | P-value | Risk of Bias assessment |
|-----------|--------------|-----------|------------|-----------------------------|-------------|------------------|------------------|---------|------------------------|
| Citalopram | 60 vs. 40 | 1 (38) | 129 vs. 131 | 0.07 [-0.17; 0.31] | 0.573 Unclear | 10 (38) | 60 vs. 40 |
| Citalopram | 60 vs. 20 | 1 (38) | 129 vs. 130 | -0.20 [-0.44; 0.05] | 0.112 Unclear | 10 (38) | 60 vs. 20 |
| Citalopram | 50 vs. 25 | 1 (61) | 8 vs. 10 | -0.39 [-0.58; 0.18] | 0.898 Unclear | 10 (61) | 50 vs. 25 |
| Citalopram | 40 vs. 20 | 1 (38) | 8 vs. 10 | 0.91 [0.60; 0.22] | 0.963 Unclear | 10 (38) | 40 vs. 20 |
| Fluvoxamine | 150 vs. 100 | 1 (63) | 48 vs. 61 | 0.12 [-0.26; 0.49] | 0.551 High | 10 (63) | 150 vs. 100 |
| Fluvoxamine | 100 vs. 50 | 1 (61) | 48 vs. 56 | 0.08 [-0.31; 0.47] | 0.682 High | 10 (61) | 100 vs. 50 |
| Fluvoxamine | 50 vs. 25 | 1 (61) | 61 vs. 65 | 0.93 [-0.43; 0.31] | 0.751 High | 10 (61) | 50 vs. 25 |
| Fluoxetine | 20 vs. 10 | 3 (13,34) | 343 vs. 334 | -0.01 [-0.21; 0.20] | 0.963 Unclear | 10 (13,34) | 20 vs. 10 |
| Fluoxetine | 15 vs. 10 | 1 (63) | 160 vs. 160 | 0.33 [0.26; 0.41] | 0.751 High | 10 (63) | 15 vs. 10 |
| Paroxetine | 40 vs. 20 | 1 (39) | 100 vs. 104 | 0.10 [-0.17; 0.38] | 0.457 Unclear | 10 (39) | 40 vs. 20 |
| Paroxetine | 30 vs. 20 | 1 (39) | 99 vs. 104 | 0.11 [-0.17; 0.38] | 0.443 Unclear | 10 (39) | 30 vs. 20 |
| Paroxetine | 20 vs. 10 | 1 (39) | 99 vs. 100 | 0.32 [-0.60; -0.04] | 0.024 Unclear | 10 (39) | 20 vs. 10 |
| Paroxetine | 10 vs. 5 | 1 (63) | 6 vs. 6 | 0.76 [-0.93; 0.41] | 0.202 Unclear | 10 (63) | 10 vs. 5 |

A negative standardized mean difference (SMD) indicates a better effect of the higher dose.

10 mg paroxetine and 12.5 mg paroxetine-controlled-release as well as 20 mg paroxetine and 25 mg treated as the same dose because paroxetine CR releases only about 80% of its active compound (40).
per arm. Other low risk of bias studies were not available.

Sensitivity NMA analyses regarding efficacy of low vs. medium vs. high doses. Twenty-two studies reporting 26 comparisons form the base of our network meta-analysis, with one closed loop (low-medium-high, Figure S2) and transitivity assumption met (Table S4). With the exception of the comparison between medium and low dosages (SMD $-0.15$ [$-0.26$; $-0.04$]), there were no comparisons with $P$-values lower than 0.05, and all effect estimates were equal to or below standardized mean differences of 0.15. Overall heterogeneity was 49.8% ($I^2$; Table S4; Figure S3).

When placebo groups were included (30 studies, 71 pairwise comparisons), results did not change, but showed medium and high SSRI superior to placebo, with effect sizes (SMDs) between 0.17 and 0.22 (Table S4).

As in our analyses of direct comparisons, we excluded studies by Guy et al. (1986) (27), Ghose et al. (1997) (35), Yevtushenko et al. (2007) as well as the studies SER101 and SER310 due to high inconsistency but results did not change substantially: With 17 studies left, all SMDs remained below $-0.10$ (data not shown).

Analysis carried out after submission of the manuscript. In contrasting all minimum recommended SSRI doses to all SSRI doses above we found a SMD of $-0.015$ [$-0.079$; 0.050] ($P = 0.652$, $I^2$: 27%), with no estimates for single substances differing substantially from the summary estimate for all SSRIs. The summary estimate of all doses below recommended SSRI doses versus all SSRI doses above amounted to a SMD of $-0.176$ [$-0.363$; 0.012], $P = 0.066$, $I^2$: 64%. In this analysis, paroxetine (SMD: $-0.297$ [$-0.473$; $-0.122$], $P = 0.001$) and, to a lesser degree, citalopram (SMD: $-0.214$ [$-0.733$; 0.305], $P = 0.420$) stood out.

Discussion

This analysis revealed three important results: (i) There is no consistent indication of statistically or clinically significant differences among SSRI doses at the level of the categories of low, medium, and high doses. (ii) In 56 published comparisons of specific doses almost no clinically relevant dose–response relationship surfaced, with few exceptions, for example, paroxetine 10 vs. 20 mg (higher better) and fluoxetine 20 vs. 60 mg (higher worse, evidence unambiguous). (iii) Relative to low doses, high and medium doses come with more dropouts due to side effects.

This meta-analysis extends the most recent pairwise meta-analyses (7,8) by more than a dozen new studies and adds to recently published meta-analyses using indirect evidence. (5,6) By categorizing doses into ranges, we added another approach to classical comparisons of specific doses. We employed several sensitivity analyses, including a network meta-analysis, and thus believe in the robustness of our findings.

We did, however, find signals of a positive dose–response gradient with paroxetine: on aggregation of four studies (9,10,12,39) 20 mg was superior to 10 mg. Results of further comparisons, such as 10 mg vs. 30 or 40 mg (39) or 15 mg vs. 30 mg (35) paroxetine, support the assumption that 10 mg paroxetine seems to be inferior to higher doses. Further, this finding is in accordance with results by Hieronymus and co-authors in an individual patient data meta-analysis (40) showing that 10 mg paroxetine (and low doses of citalopram and sertraline) seems to be inferior to higher doses, as measured by a single-item on the Hamilton Scale (depressed mood). However, 10 mg paroxetine is not recommended and, in our experience, already rarely used in clinical practice.

More conclusive for clinical decision-making is the superiority of 20 mg fluoxetine versus 60 mg (a borderline effect, as measured), suggesting it is not only useless to increase fluoxetine to 60 mg but harmful, in particular considering the increased risk for dropouts.

At the level of dose ranges and for SSRIs as a group, summary estimates did not exceed our definition of a clinically meaningful effect (SMD 0.2). The difference between medium and low doses reached a $P$-value below 0.05, but was smaller (SMD 0.15) than what is normally considered a lower threshold of a small effect. In addition, this small effect was not supported in one sensitivity analysis (Table 1). In another sensitivity analysis, carried out after submission of the manuscript, we combined all recommended SSRI doses and all SSRIs doses above in a comparison with all SSRI doses below recommended levels and found a statistically non-significant and relatively low summary SMD ($-0.176$), but with a wide confidence interval [$-0.36$; 0.01]. Together with the wide CI of our primary analysis (0.28; $-0.01$ SMD, Table 1) and in the presence of substantial heterogeneity, we consider the results inconclusive. Therefore, we believe that there is not enough reason to dose lower than what has been recommended as minimum doses, that is, lower than 10 mg escitalopram or 20 mg fluoxetine, for example. In some cases, there may be clinical reasons to do so, for example, related to age or to tolerability. However, the effect
direction, the confidence interval, and the heterogeneity in our analyses as well as some direct comparisons among single substances (Table 3) do not favor lower than recommended doses. On balance, in clinical decision-making, prudence suggests to aim for the minimum recommended dose.

From a clinical viewpoint, the comparison of medium and high doses seems more relevant, and differences were entirely negligible in both our primary pairwise and our sensitivity network meta-analyses. However, the confidence interval is so wide that clinically relevant, effect differences cannot be excluded (high vs. low doses: lower limit of $-0.25$ SMD). As regards specific doses of single SSRIs—as opposed to dose ranges—there are several broad confidence intervals, but we note that sample sizes are often small.

Our findings resonate with an earlier meta-analysis in which we showed that, in general, dose increase seems to be no promising strategy for non-responders to a standard dose of the particular antidepressant (19). The results also extend conclusions of the earlier systematic reviews and meta-analyses on the topic from more than fifteen years ago (7,8,41), and they also fit in the broader picture of dose escalation strategies in a recent meta-analysis (42). Interestingly, the results are also very close to those published by Furukawa et al. (6).

In a SPECT study, Ruhé et al. (43) point out a possible explanation for the lack of a dose–response gradient with SSRIs: Non-responders to a trial of 20 mg paroxetine were randomly assigned to continuing 20 mg or to a dose increase to a mean of 46.7 mg. At the lower dose, about 80 % of the individual’s serotonin transporters were blocked, and after dose increase, the occupancy rate stayed at 80 %. Antidepressive effects of both doses were identical.

Our results do not line up with the conclusions in the meta-analysis by Jakubovski et al. (5). An important difference is that these authors reported different dosages were associated with different results among studies but not within studies. The studies did not randomly assign patients to different SSRI dosages while the present analysis is based on direct comparisons of SSRIs—the difference between level II and level I evidence. Furthermore, the inclusion of flexible-dose arms but treating patients of these arms as if they had received the highest permissible dose might have biased the results. Finally, it is hard to see the clinical relevance of the effect size published by Jakubovski et al. (5): Regarding the continuous, and thus statistically most informative, measurement of differences in depression scores, there was an increase of approximately $0.08$ SMD over the range of 100 to 250 mg imipramine equivalents, a very small effect by any standard (22,44). Interestingly, the effect size is similar to several of the effect sizes found in our analyzes.

Montgomery et al. (45) concluded from an analysis of nine citalopram studies that 40 mg is better than 20 mg for severe or recurrent depression. It seems the authors did not directly compare doses but different doses relative to placebo, thereby increasing the danger of type I errors. In our study, together with paroxetine and fluoxetine, citalopram turned out to be the most intensely studied compound, and we could not find strong evidence for different efficacy under different doses.

Neither our meta-analysis nor the studies we analyzed refer to serum concentrations of SSRIs. However, for most SSRIs a therapeutic range is established (46), and it is of clinical importance to note that for individuals with a sub-therapeutic serum concentration despite standard-dose treatment (e.g., due to rapid metabolism), a dose increase can be a reasonable therapeutic strategy.

Even though we did not find solid evidence for a dose–response relationship regarding antidepressant efficacy, dropouts, in particular caused by adverse events, clearly rise with increasing dose. The effect sizes of the differences in dropouts due to side effects are in the small to medium range. In addition, in contrast to the effect sizes of antidepressant efficacy calculated in this analysis, $P$-values are consistently below 0.05 and thus represent a consistent finding. In our view, the increasing number of side effect related dropouts with dose is a very important additional reason for prudence in considering a dose increase beyond medium levels in SSRI treatment. Of note, fixed-dose studies run the risk of higher dropouts due side effects as clinicians cannot adapt doses according to their tolerability.

In dose increase scenarios, decisions are often based on a trade-off between effectiveness and tolerability. This is why we restricted our notion to routine treatment: In individual cases, there can be valid reasons to favor a high dose. It is also worth noting that there are some data indicating a dose–response relationship for other groups of antidepressants, particularly tricyclic antidepressants (41).

Limitations

As in all systematic reviews we may have missed important studies: It is imaginable that dose–effect differentials surfaced in very early, unpublished investigations, for example, in phase-II-trials. However, while not all studies have been published, and hence could not have been included in our meta-
analysis, it is likely that such a publication bias would work in the direction of our finding: It is in the interest of drug manufacturers and researchers alike to publish positive studies on dose–response effects (inasmuch as the relationship was positive). We may also have missed published studies, but our literature search was systematic and extensive, and followed the approach of the Cochrane collaboration.

The recommendations by manufacturers may rest in part on early dose-finding studies, and, in turn, our operationalization of low, medium, and high doses reflect these recommendations. Still, the field keeps investigating a dose–response gradient in both RCTs and meta-analyses (e.g., Furukawa et al. (6), Jakubovski et al. (5) and Hieronymus et al. (40)). Against this backdrop, our results not only support the recommendations by drug manufacturers but also shed light on all other dose differences investigated.

Studies using LOCF (last observation carried forward) to deal with dropouts might put dose increase at a disadvantage because dose increase may lead to dropouts during the first week and thus high depression scores are carried forward.

We would have liked to use regression slopes for diverse doses and effects, as stated in our study protocol. However, such an approach rests on an approximately linear relationship between doses and effects, and we could not find any indication for linearity (or curvilinear associations). Therefore, we modified the analysis to what we describe in the methods section. We acknowledge the limits to our joint analysis of SSRIs employing high, medium, and low doses based on product information. For example, recommendations may vary according to region or time. However, while there seems to be no perfect method to classify dose ranges across SSRIs and while the method chosen leads to a preponderance of medium doses, the one selected for our analysis provides external, transparent, and universally accepted dose ranges.

We relied on placebo response estimates for a sensitivity network meta-analysis, but there is debate about placebo response rates varying across the observation period of our study (47,48). A recent comprehensive systematic review demonstrated stability of placebo response in antidepressant trials over the last 25 years (47).

The quality of reporting was moderate, as is often the case with RCTs published before reporting guidelines came into use, and we needed to impute several standard deviations. In our experience, however, this is not unusual in meta-analyses of older studies and we followed general rules for imputation. The reporting quality of many of the older studies and the risk of bias throughout the study pool are a caveat to the robustness of our results, as is the fact that several comparisons are underpowered. In a similar vein, we noted statistical heterogeneity of results in quite a few of our analyses. Also, as always in RCTs, we cannot be sure that blinding has always worked as intended (49). Interestingly, study quality seems not to be closely linked to treatment estimates (50). In any case, these limitations strengthen our main finding that the notion of a dose–response relationship in SSRIs is not well supported by data.

Meta-analyses, like other scientific methods, have general limitations. Specifically, they are an attempt at shrinking an abundance of data from a variety of studies to digestible amounts of information. Any such simplification runs the risk of being too simple, for example, because studies are methodologically, conceptually different. This is particularly problematic in meta-analyses employing indirect comparators, such as network meta-analysis, and suggests the use of direct comparisons as in this study.

In total, the picture seems quite homogeneous in indicating that there are no clinically relevant dose–effects gradients of SSRIs. At the very least, it is fair to say that we lack evidence for such a dose–response relationship. It has become clear, however, that higher doses come with the risk of more dropouts.

Acknowledgement

Open access funding enabled and organized by Projekt DEAL.

Funding

This study has not been externally funded.

Conflict of interest

The authors declare that they have no conflicts of interest with regard to the topic of this manuscript.

Peer Review

The peer review history for this article is available at https://publons.com/publon/10.1111/acps.13235.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Loth M, Müller-Oerlinghausen B. Psychopharmacol. In: Schwabe U, Paffrath D, Ludwig W-D, Klauber J, editors.
14. Hirayasu Y, Gomossil C, Chen D, Nunez R, Khan A. Efficacy and safety of vilazodone 20 and 40 mg in major depressive disorder: a randomized, double-blind, placebo-controlled trial. Int Clin Psychopharmacol 2015;30:67–74.

15. Mathews M, Gomossil C, Chen D, Nunez R, Khan A. Efficacy and safety of vilazodone 20 and 40 mg in major depressive disorder: a randomized, double-blind, placebo-controlled trial. Int Clin Psychopharmacol 2015;30:67–74.

16. Kato T, Furukawa TA, Mantani A et al. Optimising first- and second-line treatment strategies for untreated major depressive disorder - the SUND study: a pragmatic, multi-centre, assessor-blinded randomised controlled trial. BMC Med 2018;16:103.

17. FDA. Selective Serotonin Reuptake Inhibitors (SSRIs) Information; 2014. https://www.fda.gov/drugs/information-drug-class/selective-serotonin-reuptake-inhibitors-ssris-information. Accessed 29.06.2020.

18. Higgins J, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.hanbook.cochrane.org

19. Rick L, Braun C, Bischor T, Henssler J, Franklin J, Baetig C. Dose increase versus unchanged continuation of antidepressants after initial antidepressant treatment failure in patients with major depressive disorders: a systematic review and meta-analysis of randomized, double-blind trials. J Clin Psychiatry. 2018;79:17r11693.

20. Licht RW, Quitzau S. Treatment strategies in patients with major depression not responding to first-line serotonin reuptake treatment. A randomised study of extended duration of treatment, dose increase or mianserin augmentation. Psychopharmacology 2002;161:143–151.

21. Henssler J, Kirschus M, Franklin J, Bischor T, Baetig C. Long-term acute-phase treatment with antidepressants, 8 weeks and beyond: a systematic review and meta-analysis of randomized, placebo-controlled trials. J Clin Psychiatry. 2018;79:15r10545.

22. Cohen J. Statistical power analysis for the behavioral sciences, 2nd edn. Lawrence Erlbaum Associates; 1988: 25.

23. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629–634.

24. Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting biases. In: Higgins JP, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: handbook.cochrane.org

25. Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. J Clin Epidemiol. 2006;59:7–10.

26. FDA. U.S. Food and Drug Administration. Medication guide (sertraline hydrochloride), Reference ID: 4192847; 2017. https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/019839s091lbl.pdf#page=27. Accessed 10.07.2018.

27. Guy W, Manov G, Wilson WH. Double-blind dose determination study of a new antidepressant—sertraline. Drug Dev Res 1986;9:267–272.

28. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997;50:683–691.

29. Rücker G. Network meta-analysis, electrical networks and graph theory. Res Synth Methods 2012;3:312–324.

30. CINeMA: Confidence in Network Meta-Analysis [Software]. University of Bern; 2017. Available from: cinema.ispm.ch [computer program].

31. Schwarzer G, Carpenter JR, Rücker G. Network meta-analysis. In: Schwarzer G, Carpenter JR, Rücker G (eds.), Meta-Analysis with R. Cham: Springer; 2015: 187–216.
randomized controlled trials and treatment estimates: systematic review, meta-analysis, and meta-regression analysis. J Clin Psychopharmacol 2007;27:352–356.

41. BHERIENSTEIN H, EDMAN G, FLUCK L, HAGENFELDT L, SEVALL G, WIELS FA. Clinical and biochemical effects of citalopram, a selective 5-HT reuptake inhibitor—a dose-response study in depressed patients. Psychopharmacology 1985;87:253–259.

42. Study 89306 (FDA). Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/98/020822a_medr_P2.pdf

43. SCT-MD-49 (NCT00668525). Available from: https://clinicaltrials.gov/ct2/show/NCT00668525

44. WERNICKE JF, DUNLOP SR, DORSENF BE, ZERBE RL. Fixed-dose fluoxetine therapy for depression. Psychopharmacol Bull 1987;23:164–168.

45. DUNLOP SR, DORSENF BE, WERNICKE JF, PUTVIN JH. Pattern analysis shows beneficial effect of fluoxetine treatment in mild depression. Psychopharmacol Bull 1990;26:173–180.

46. FABRE LF, PUTMAN HP 3rd. A fixed-dose clinical trial of fluoxetine in outpatients with major depression. J Clin Psychiatry 1987;48:406–408.

47. FIEVE RR, GOODNICK PJ, PSELEOW E, SCHLEGEL A. Fluoxetine response: endpoint vs pattern analysis. Int Clin Psychopharmacol 1986;1:320–323.

48. WERNICKE JF, DUNLOP SR, DORSENF BE, BOSOMWORTH JC, HUMBERT M. Low-dose fluoxetine therapy for depression. Psychopharmacol Bull 1988;24:183–188.

49. TURAI, ONDER E. Fluoxetine once every third day in the treatment of major depressive disorder. Eur Arch Psychiatry Clin Neurosci 2003;253:307–312.

50. NCT01808612. Available from: https://clinicaltrials.gov/ct2/show/NCT01808612

51. Walmack DD, APTER JT, Halkias JA et al. The oral dose-effect relationship for fluvoxamine: A fixed-dose comparison against placebo in depressed outpatients. Ann Clin Psychiatry 1996;8:139–151.

52. SURI RA, ALTHSULER LL, RASGON NL et al. Efficacy and response time to sertraline versus fluoxetine in the treatment of unipolar major depressive disorder. J Clin Psychiatry 2000;61:942–946.

53. (SB 659746–003); Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/022567Orig1s000MedR.pdf

54. ROSENBERG C, DAMBRO N, FUELUM E, JACOBSEN LV, HORGÅRD S. Citalopram and imipramine in the treatment of depressive patients in general practice. A Nordic multicentre clinical study. Int Clin Psychopharmacol 1994;9(Suppl 1):41–48.

55. HEBENSTREIT GF, FELLERER K, ZÖCHLING R, ZENTZ A, DUNBAR GC. A pharmacokinetic dose titration study in adult and elderly depressed patients. Acta Psychiatr Scand Suppl 1989;350:81–84.

56. (EMD 68 843–010). Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/022567Orig1s000MedR.pdf

**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

- Figure S1. Funnel plot.
- Figure S2. Network plot.
Figure S3. Forest plot, network meta-analysis.
Table S1. Search history.
Table S2. Dose-classification.
Table S3. Characteristics of studies.

Table S4. Network meta-analysis: primary outcome (SSRI doses categorized as low, medium, high) and sensitivity analysis including placebo arms.
Table S5. Risk of bias analysis.