Is Escitalopram Really Relevantly Superior to Citalopram in Treatment of Major Depressive Disorder? A Meta-analysis of Head-to-head Randomized Trials

Aim To evaluate clinical relevance of differences between escitalopram and citalopram (equimolar) for major depressive disorder.

Methods Review and meta-analysis of comparative randomized controlled trials (RCT). Comparisons were in relation to Montgomery-Asberg depression rating scale (MADRS) score reduction at weeks 1 (5 RCTs), 4 (5 RCTs), 6 (4 RCTs), 8 (5 RCTs), and 24 (1 RCT); proportion of responders at weeks 2, 4, 6 (2 RCTs for each time point), 8 (5 RCTs), and 24 (1 RCT); clinical global impression-severity (CGI-S) reduction at weeks 6 (1 RCT), 8 (5 RCTs), and 24 (1 RCT), and discontinuation due to adverse events or inefficacy during short-term (up to 8 weeks) and medium-term (24 weeks) treatment.

Results MADRS reduction was greater with escitalopram, but 95% confidence intervals (CI) around the mean difference were entirely or largely below 2 scale points (minimally important difference) and CI around the effect size (ES) was below 0.32 (“small”) at all time points. Risk of response was higher with escitalopram at week 8 (relative risk, 1.14; 95% CI, 1.04 to 1.26) but number needed to treat was 14 (95% CI, 7 to 111). All 95% CIs around the mean difference and ES of CGI-S reduction at week 8 were below 0.32 points and the limit of “small,” respectively. Data for severe patients (MADRS ≥ 30) are scarce (only 1 RCT), indicating somewhat greater efficacy (response rate and MADRS reduction at week 8, but not CGI-S reduction) of escitalopram, but without compelling evidence of clinically relevant differences. Discontinuations due to adverse events or inefficacy up to 8 weeks of treatment were comparable. Data for the period up to 24 weeks are scarce and inconclusive.

Conclusion Presently, the claims about clinically relevant superiority of escitalopram over citalopram in short-to-medium term treatment of major depressive disorder are not supported by evidence.
Due to its common occurrence and burden that it represents for the affected individuals and society, depression is a paramount health care problem (1). Understandably, treatment options have attracted much attention and any new treatment showing potential for improvements is of a considerable interest.

Citalopram is a generic name for an active compound which is a racemic mixture (1:1) of two enantiomers \( S^{(+)} \) and \( R^{(-)} \) (2). It is a selective serotonin re-uptake inhibitor (immediate-release tablets) approved as effective and safe in treatment of major depressive disorder (MDD) (3). Re-uptake inhibition and clinical effects are ascribable to the \( S^{(+)} \)-enantiomer (active enantiomer) (2). Escitalopram is a generic name for an active compound containing only the \( S^{(+)} \)-enantiomer. It was developed (after citalopram) and is marketed as effective and safe treatment for MDD by the same company that developed citalopram (4). In line with the fact that doses of citalopram and escitalopram, equimolar by content of the active enantiomer, are equivalent in terms of its in vivo bioavailability (5), the effective doses of citalopram for MDD have two times greater mass than the recommended doses of escitalopram (eg, 20 mg or 40 mg/d vs 10 mg or 20 mg/d) (3,4).

A number of post-hoc assessments have been released over the last 6 years that deal with comparative efficacy/safety of equimolar (active enantiomer) doses of escitalopram vs citalopram in MDD. They all refer to some or all of the same 6 randomized controlled trials: 5 published (6-10) (the first one in 2002 [6], last one in 2007 [10]) and one unpublished trial submitted to the Food and Drug Administration (code SCT-MD-02) (11). All trials were sponsored by the manufacturer. Some of these assessments, authored/co-authored by the company’s employees (2,12-15,) or receiving support by the manufacturer (16,) have extensively emphasized numerical trends or significant differences in some of the trials in favor of escitalopram.

Some independent reviews (eg 17,) have also emphasized that, in comparison with most of the other newer antidepressants (including citalopram), escitalopram provides relevantly better effects. Consequently, escitalopram has been rather aggressively marketed as a superior treatment of depression (18) with a resulting remarkable sales growth paralleled by a drop in citalopram use over the last 6 years that deal with comparative efficacy/safety of equimolar (active enantiomer) doses of escitalopram vs citalopram in MDD. They all refer to some or all of the same 6 randomized controlled trials: 5 published (6-10) (the first one in 2002 [6], last one in 2007 [10]) and one unpublished trial submitted to the Food and Drug Administration (code SCT-MD-02) (11). All trials were sponsored by the manufacturer. Some of these assessments, authored/co-authored by the company’s employees (2,12-15,) or receiving support by the manufacturer (16,) have extensively emphasized numerical trends or significant differences in some of the trials in favor of escitalopram. Some independent reviews (eg 17,) have also emphasized that, in comparison with most of the other newer antidepressants (including citalopram), escitalopram provides relevantly better effects. Consequently, escitalopram has been rather aggressively marketed as a superior treatment of depression (18) with a resulting remarkable sales growth paralleled by a drop in citalopram use over the past seven years worldwide (19). The clinical observations were ascribed to the fact that the \( R^{(-)} \)-enantiomer was not just an inactive by-stander, but rather that it interfered with the activity of the \( S^{(+)} \)-enantiomer (2). On the other hand, several independent analytical assessments, although recognizing numerical trends or statistical significance of certain differences, remained restrained regarding practical relevance of escitalopram vs citalopram differences (eg 20,22,). Moreover, a recent direct-indirect (“network”) meta-analysis found no significant difference between the two drugs regarding response rates (efficacy) or discontinuation rates (tolerability) (21). Other authors (23) have explicitly pointed-out certain inconsistency in the observations/claims about the escitalopram-citalopram differences, such as superior efficacy of escitalopram vs citalopram in “moderately ill” but not “severely ill” patients (2,89 or just the opposite (12,13,15), and/or concluded that the differences between the two drugs were below practical relevance (23-27).

With such contradictory messages and with emerging awareness that we have been overestimating the efficacy of antidepressants (28-30), it seemed plausible to re-visit the trial data and try to estimate the size and clinical relevance of differences between escitalopram and citalopram in MDD with a focus on contrasts between equimolar doses and with regard to duration of treatment and initial disease severity.

METHODS

The present study was conceived as an analysis (and meta-analysis, where appropriate) of randomized controlled trials (RCT) directly comparing escitalopram and citalopram in depression. No indirect or combined direct-indirect comparisons between treatments were intended.

Literature search, trial selection, and quality assessment

The Cochrane review and meta-analysis of escitalopram for depression published in 2009 (22) was used to trace the 6 trials that it embraced. The list of excluded studies was also checked. Reference lists of other recent reviews/meta-analyses (14,15,17,21,24-27) were checked for information on potential additional trials of interest. Data on one unpublished trial (SCT-MD-02) were retrieved from the Food and Drug Administration assessment of escitalopram (11). This document provided additional information on 2 published trials (6,7).

Searching Scopus and PubMed (October 2009) (“escitalopram” [title, abstract, keywords]) and “randomized controlled trial” [publication type]) retrieved a total of 256 and 437 citations, respectively; a search of the Cochrane Central Register of Controlled Trials (November 2009) (“escitalopram” [title, abstract, keywords]) retrieved a total of 437 citations, respectively; a search of the Cochrane Central Register of Controlled Trials (November 2009) (“escitalopram” [title, abstract, keywords]) retrieved a total of 437 citations, respectively; a search of the Cochrane Central Register of Controlled Trials (November 2009) (“escitalopram” [title, abstract, keywords]).
tualpam AND citalopram” [title, abstract, keywords] AND “randomized controlled trial” [publication type]) retrieved 114 citations; a search of the EBSCO Publishing Electronic Databases (November 2009) “escitalopram” AND “citalo-

The 6 trials embraced by the Cochrane group meta-analysis (22) were all judged to be of fair quality by the Cochrane collaboration criteria (22) and were not further evaluated in this respect. Any additional trial that was to be included in the current analysis was to be assessed for quality by the same criteria (31).

Outcomes of interest and data extraction

The current regulatory requirements (32) for antidepressants are that shorter trials for initial treatment should have 2 primary outcomes for assessment of efficacy, both based on a validated depression severity measuring scale (Montgomery-Asberg depression rating scale [MADRS], Hamilton depression rating scale [preferably the 17-item version, HAM-D] (33,34); reduction of symptom severity vs baseline (a continuous variable) and a proportion of patients experiencing ≥50% reduction in disease severity (proportion of responders) at endpoint. Changes in Clinical Global Impression of Disease Severity (CGI-S vs baseline) (35), although less specific for depressive symptoms than the abovementioned scales, provide useful information on patient clinical status (36) and are suggested (32) as a secondary indicator of efficacy. All efficacy analyses should employ the “intent-to-treat” and “last observation carried forward” principles (ITT-LOCF). Hence, the 3 measures of efficacy (ITT-LOCF) were planned to be evaluated in the current escitalopram vs citalopram comparison: a) mean change in symptom score as assessed by MADRS (or HAM-D) vs baseline; b) response rates, ie, proportion of patients achieving ≥50% reduction in MADRS (or HAM-D) vs baseline; and c) mean change in CGI-S score vs baseline. Where both MADRS and HAM-D were available, MADRS was preferred and HAM-D data were considered only if MADRS data were missing.

Since safety profiles of escitalopram and citalopram in MDD (in general and in comparison) are well-known, safety assessment focused on discontinuation rates due to adverse events (AE) or lack of efficacy (or either). Safety population (all randomized patients who received at least one dose of study drug) served as a denominator.

Where data were presented only in graphical format, Dagga software (Blue Leaf Software, Hamilton, New Zealand) that digitalizes drawings (in PDF or any other format) was used to recover numerical values.

Where standard deviations (SD) for “mean change vs baseline” (MADRS, HAM-D, CGI-S) were missing, the preferable strategy was to recover them using the Cochrane methodology (31) from any of the following (if available): standard errors, confidence intervals, test statistic, or P-values. Where no recovery was possible, SDs were imputed using either of 2 validated methods (37): imputation of pooled SD of all other treatment arms in the remaining included studies or imputation of SD “borrowed” from other trials (similar in design) or other meta-analysis. With SD imputation, care was taken to use the values that corresponded to the “flawed” report in treatment duration (eg, if SD was missing for a change at week 8, then the imputed SD had to reflect the SD of change at week 8). When there were multiple flaws in data reporting requiring several approximations (eg, low-quality graphical data on mean change with no information on variability), trial data were not used for this particular outcome. Where only mean between-treatment differences were reported but not actual treatment means, pooled SD was recovered (as described) and used to determine effect size for use in meta-analysis.

Data evaluation and synthesis

Where data from more than one trial were available, random effects meta-analysis of the efficacy and safety outcomes was conducted using StatsDirect software, version 2.7.7. (StatsDirect Ltd, Altrincham, UK) that is based on the methodology of the Cochrane collaboration and was employed in some of the meta-analysis cited here (eg 24-26, ). All estimates of between-treatment differences are reported with 95% confidence intervals (CI). Differences in “mean changes vs baseline” are expressed as mean difference (single trial) or weighted mean difference (WMD, from meta-analysis) and also as a standardized mean difference (SMD) (wherever possible). The criteria for assessment of clinical relevance were: a) for MADRS, 2 scale points as a limit of a relevant difference (mean, WMD) in reduction since it is considered as “minimally important difference” (17); b) for CGI-S, a difference between treatments in change vs baseline (mean or WMD)≥0.33 scale points was arbitrarily chosen as practically relevant (ie, differences ≤0.32 are practically irrelevant); c) for SMD (any instrument), the value of 0.32 as a limit of “small effect size” (ES) was used as a limit of relevance: practicallyclinically relevant if
A total of 7 trials were included, ie, 1 additional (39) (Table 1) as compared with the most recent meta-analysis published in 2009 (22). The additional trial (39) was also industry-sponsored, but not by the originator of citalopram and escitalopram. Its quality is limited mainly due to selective reporting: no numerical data for post-baseline depression scale (HAM-D$_{17}$) and CGI-S scores were provided, no measures of variability were reported, and graphical presentation of means did not allow for a reasonably reliable data extraction (39).

All trials were double-blind, parallel group multicentric RCTs conducted at different geographical locations and included adult, otherwise healthy, and mainly younger out-patients with MDD free of other psychopathology (Table 1). They differed in duration (minimum 4, maximum 24 weeks) (Table 1) and primary objectives (superiority of escitalopram vs placebo, non-inferiority or superiority vs citalopram) (Table 1). Fixed doses of either drug (no dose-adjustment) were delivered in 4 trials, whereas dosing was flexible (initial doses could be doubled depending on response/tolerability) in 3 trials (Table 1). Of those, average delivered doses were reported in 2 trials (7,11) and could be judged as equimolar, whereas in the third trial (39) the proportion of patients with increased doses was roughly comparable for the escitalopram and citalopram treatment arms (Table 1). Five trials were multi-arm (placebo and/or an additional dose of either drug or an additional active treatment), but the current evaluation focused on equimolar escitalopram vs citalopram treatment arms (Table 1). In all trials, efficacy was assessed based on the intent-to-treat (ITT, all patients receiving at least one dose of study drug and having at least one post-baseline evaluation) data set with employment of the last observation carried forward principle, whereas safety assessment considered all patients receiving at least one dose of study drug (Table 1). Depression was quantified using MADRS in 6 trials and using HAM-D$_{17}$ in 1 trial (Table 1). Each trial provided information about at least one efficacy outcome on at least one additional time point besides the endpoint (Table 2).

**Efficacy comparison of escitalopram and citalopram**

Figure 1 summarizes escitalopram vs citalopram differences in MADRS reduction and response rates by evaluation time point. Web extra material 1 (MADRS reduction) and web extra material 2 (response rates) provide details on individual trial data and conducted meta-analysis. Figure 2 shows individual trial data and meta-analysis of CGI-S change vs baseline at week 8.

**MADRS change vs baseline**

MADRS reduction with escitalopram was consistently greater than with citalopram at weeks 1, 4, 6, 8, and 24 (Figure 1). However, 95% CI around WMD was entirely or almost entirely below 2 points at all times except week
6 (largely below 2 points), and 95% CI around SMD was practically entirely below 0.32 at all time points (Figure 1). Confidence intervals around mean or standardized mean between-treatment differences at weeks 1, 4, and 6 were entirely or largely below the limits of “relevance” for all individual trials except for one (10) (web extra material 1). At week 6 (endpoint), point-estimates in this particular trial were largely in the range of “practically relevant” and were the source of mild to moderate inconsistency (web extra material 1). At week 8, 95% CI around all individual standardized between-treatment differences were entirely or largely below the limit of “relevance” (web extra material 1). Only one trial yielded a mean difference point estimate larger than 2 scale points (-2.1 in the trial by Moore et al [9], borderline significant) (web extra material 1).

The 4-week trial based on HAM-D, (39) and not included in this meta-analysis (no numerical data reported, no measures of variability) provided a graphical representation of changes vs baseline at weeks 1, 2, 3, and 4 with complete-

<table>
<thead>
<tr>
<th>Author (reference)</th>
<th>Design/setting</th>
<th>Aim</th>
<th>Patients</th>
<th>Treatment arms (mg/d)</th>
<th>Baseline MADRS (ITT) (mean±SD)</th>
<th>Dosing (mg/d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burke (6)</td>
<td>parallel DB, RCT MC, USA 1 week placebo run-in 8 weeks treatment psychiatrists</td>
<td>Superiority ESC vs placebo</td>
<td>Otherwise healthy MDD outpatients (DSM-IV), MADRS≥22, age 18-65, no other psychopathology</td>
<td>placebo: n = 119/122 ESC 10: n = 118/119 ESC 20: n = 123/125 CIT 40: n = 125/125</td>
<td>29.5 ± 5.0 28.0 ± 4.9 29.2 ± 4.6 29.2 ± 4.5</td>
<td>Fixed 1st week 10, then 20 1st week 20, then 40</td>
</tr>
<tr>
<td>Lepola (7)</td>
<td>parallel DB, RCT MC, Canada &amp; EU 1 week placebo run-in 8 weeks treatment primary care</td>
<td>Superiority ESC vs placebo</td>
<td>Otherwise healthy MDD outpatients (DSM-IV), MADRS 22-40, age 18-65, no other psychopathology</td>
<td>placebo: n = 154/154 ESC 10-20: n = 155/155 CIT 20-40: n = 159 / 160</td>
<td>28.7 ± 4.0 29.0 ± 4.3 29.2 ± 4.2</td>
<td>Flexible – ESC 10, CIT 20 could double at week 4/6 Mean: 14.0 Mean: 28.4</td>
</tr>
<tr>
<td>SCT-MD-02 (11)</td>
<td>parallel DB, RCT MC, USA 1 week placebo run-in 8 weeks treatment psychiatrists</td>
<td>Superiority ESC vs placebo</td>
<td>Otherwise healthy MDD outpatients (DSM-IV), MADRS≥22, age 18-80, no other psychopathology</td>
<td>placebo: n = 125 / 127 ESC 10-20: n = 124 / 125 CIT 20-40: n = 119 / 123</td>
<td>28.8 ± 5.0 28.7 ± 4.3 28.3 ± 5.0</td>
<td>Flexible – ESC 10, CIT 20 could double at week 3-8 Mean: 17.6 Mean: 35.3</td>
</tr>
<tr>
<td>Lalit (39)</td>
<td>parallel DB, RCT MC, India placebo run-in 4 weeks treatment psychiatrists</td>
<td>Not clearly defined Likely inequality ESC/CIT/SER</td>
<td>Otherwise healthy MDD outpatients (ICD-10, HAM-D≥18, age 18-65, no other psychopathology</td>
<td>ESC 10-20: n = 69/69 CIT 20-40: n = 74/74 SER 100: n = 71/71</td>
<td>HAM-D17, 26 ± 6 25 ± 5 25 ± 4</td>
<td>Flexible – ESC 10, CIT 20 could double at week 2 42% at 20 mg 51% at 40 mg</td>
</tr>
<tr>
<td>Colonna (8)</td>
<td>parallel DB, RCT MC, EU 1 week placebo run-in 24 weeks treatment primary care</td>
<td>Non-inferiority ESC vs CIT</td>
<td>Otherwise healthy MDD outpatients (DSM-IV), MADRS≥22, age 18-65, no other psychopathology</td>
<td>ESC 10: n = 165/175 CIT 20: n = 174/182</td>
<td>29.8±not reported 29.8±not reported</td>
<td>Fixed</td>
</tr>
<tr>
<td>Moore (9)</td>
<td>parallel DB, RCT MC, France 8 weeks treatment psychiatrists and primary care</td>
<td>Superiority ESC vs CIT</td>
<td>Otherwise healthy MDD outpatients (DSM-IV), MADRS≥30, age 18-65, no other psychopathology</td>
<td>ESC 20: n = 138 142 CIT 40: n = 142/152</td>
<td>36.3 ± 4.8 35.7 ± 4.4</td>
<td>Fixed 1st week 10, then 20 1st week 20, then 40</td>
</tr>
<tr>
<td>Yevtushenko (10)</td>
<td>Parallel DB, RCT MC, Russia 6 weeks treatment psychiatrists</td>
<td>Superiority ESC vs CIT</td>
<td>Otherwise healthy MDD outpatients (DSM-IV), MADRS≥25, age 25-45, no other psychopathology</td>
<td>ESC 10: n = 108/108 CIT 20: n = 109/109</td>
<td>34.8 ± 3.5 34.3 ± 3.3 34.7 ± 3.8</td>
<td>Fixed</td>
</tr>
</tbody>
</table>

*Abbreviations: ITT – intent-to-treat; LOCF – last observation carried forward; MADRS – Montgomery-Asberg Depression Rating Scale (33); DB – double-blind; RCT – randomized controlled trial; MC – multicentric; USA – United States of America; DSM-IV – Diagnostic and Statistical Manual of Mental Disorders, 4th edition (40); EU – European Union; SD – standard deviation; SER – sertralin; ICD-10 – International Classification of Diseases and Related Health Problems, 10th revision (41); HAM-D17 – 17-item Hamilton Depression Rating Scale (34).

†Comparison of interest for the current evaluation (equimolar doses).
ly overlapping mean values for escitalopram and citalopram.

**Response rates**

Response rates were based on MADRS in all trials except one (HAM-D-17). Apart for week 8, data are scarce (Figure 1). At week 8, with mild inconsistency, “risk” of response was significantly higher with escitalopram (pooled RR around 1.14) (Figure 1). However, NNTB was below the limit of “practical relevance” (Figure 1). Only one trial (Moore et al [9]) out of 5 yielded a significant RR and an NNTB approaching “clinical relevance” (web extra material 2). Pooled estimates at week 2, 4, and 6, each based on 2 RCTs, are difficult to interpret. As shown in web extra material 2, at each time point, one of the trials yielded response rates that were for both drugs 2 to 4 times lower than in the other considered trial. Proportion meta-analysis indicates non-combinability of such proportions for either drug at any of these time points (P < 0.001 based on Cochran Q). Therefore, from the practical standpoint it appears more informative to consider each trial separately based on NNTB: at week 2, both trials indicate no relevant difference; at week 4, both trials indicate no relevant difference; at week 6, one trial indicates no relevant difference, whereas the other one indicates a difference “approaching relevance” (web extra material 2).

**CGI-S change vs baseline**

At week 8, with mild inconsistency, the 95% CIs around pooled WMD and SMD were entirely well below the limits of practical relevance (Figure 2). Also, mean and standardized between-treatment differences were entirely or largely below the limits of relevance for each individual trial (Figure 2).

### TABLE 2. Availability of information on efficacy outcomes based on intent-to-treat with last observation carried forward data by trial*

<table>
<thead>
<tr>
<th>Author (reference)</th>
<th>Burke (6)</th>
<th>Lepola (7)</th>
<th>SCT-MD-02 (11)</th>
<th>Lalit (39)</th>
<th>Colonna (8)</th>
<th>Moore (9)</th>
<th>Yevtushenko (10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used data:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>published</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>unpublished</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Duration (weeks)</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>24</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>MADRS change:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week 1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>week 4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>week 6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>week 8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>week 24</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Response rates:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week 2</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>week 4</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>week 6</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>week 8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>week 24</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>CGI-S change:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week 6</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>week 8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>week 24</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

*Abbreviations: MADRS – Montgomery-Asberg Depression Rating Scale (33); SD – standard deviation; SE – standard error; CI – confidence interval; CGI-S – Clinical Global Impression-severity (35). HAM-D<sub>17</sub> – 17-item Hamilton Depression Rating Scale (HAM-D<sub>17</sub>) (34) was used for quantification of depression but, just as with CGI-S, no post-baseline numerical data and measures of variability were reported. Graphical presentations did not allow for data extraction. Response was defined as ≥50% reduction in HAM-D<sub>17</sub>. #For week 24 SD recovered from CI of difference, for week 8 imputed (37). §Numerical values reported for week 6, for weeks 1 and 4 mean and SE graph-read. †For week 8 SD imputed, for week 24 from P.
Summary of escitalopram (ESC) vs citalopram (CIT) differences in reduction of symptom severity (upper panel) and response rates (lower panel) at different evaluation time points (weeks of treatment). Upper left panel shows weighted mean difference (WMD) with a shaded area indicating the limit of “minimally important difference,” upper right panel shows standardized mean difference (SMD) with a shaded area indicating the limit of “small” effect size. Abbreviations: MADRS – Montgomery-Asberg Depression Rating Scale (33); RCT – randomized controlled trial; RR – relative risk; NNT – number needed to treat. Asterisk – MADRS was used in all trials except one (39) that used hAM-d (34) and provided data on response rates at week 2 and week 4.

Meta-analysis of escitalopram (ESC) vs citalopram (CIT) differences in reduction of Clinical Global Impression-severity (CGI-S) (35) at week 8 vs baseline. Left panel shows weighted mean difference (WMD) with a shaded area indicating the limits of “practically relevant difference.” Right panel shows standardized mean difference indicating the limit of “small effect size.”
The 4-week trial not reporting numerical values on CGI-S changes (39) includes a graphical representation of CGI-S change vs baseline at weeks 1, 2, 3, and 4 with completely overlapping values for escitalopram and citalopram.

One trial (total N = 216) reporting on week 6 (10) indicated a large difference in favor or escitalopram (mean, -0.55; 95% CI, -0.83 to -0.27; standardized, -0.53; 95% CI, -0.8 to -0.26).

One trial (total N = 339) reporting on week 24 (8), indicated a difference in favor of escitalopram, but largely below the limit of relevance (mean, -0.24; 95% CI, -0.55 to 0.07; standardized, -0.17; 95% CI, -0.38 to 0.05).

**Efficacy comparison of escitalopram and citalopram in patients with baseline MADRS≥30**

Only one RCT was conducted specifically in patients with baseline MADRS≥30 (“severely depressed”) and was included in the main meta-analysis (Moore et al [9]). The main efficacy outcomes (Table 3) indicate a difference in favor of escitalopram that could be judged as “borderline practically relevant.” Differences in MADRS (week 4, week 8) and CGI-S (week 8) reduction are largely below the limit of relevance, but the responder rates (week 8) are higher with escitalopram with an NNTB “close to” clinical relevance.

One of the trials included in the main meta-analysis (Colonna et al [8]) provided separate data for a subgroup of severely depressed patients (finding practically identical results for all outcomes for the 2 drugs). An additional publication (13) reported on a post-hoc pooled analysis of patients with MADRS≥30 selected from 3 other trials included in the present meta-analysis (6,7,11). None of these trials implemented stratified randomization regarding disease severity. Table 3 summarizes efficacy outcomes for pooled data from these 2 reports. All differences are in favor of escitalopram, but 95% CI around differences in MADRS and CGI-S reduction, as well as NNTB (response), are largely below the limits of practical (clinical) relevance.

**Discontinuations due to AE or inefficacy or either**

With mild inconsistency, the risk of discontinuation of treatment due to AE or inefficacy during the initial period of up to 8 weeks was slightly lower with escitalopram (Figure 3). The individual trial data vary from more than twice greater risk to more than twice lower risk with escitalopram (Figure 3). Discontinuations due to AE were infrequent with both drugs with practically no difference between them (Figure 3). Discontinuations due to inefficacy were twice more frequent with citalopram than with escitalopram, but the absolute numbers are very low, ie, 9/924 (0.97%) vs 4/899 (0.44%) and the difference could have been by chance (Figure 3).

Data for the period after week 8 come only from patients completing the first 8 weeks of treatment in the only 24-week trial (8) (Figure 3). Discontinuations due to AE were apparently somewhat more frequent with citalopram, but the numbers were very low and the observed differences could have been by chance (Figure 3).

### Table 3. Summarized efficacy differences (ITT-LOCF) between escitalopram (ESC) and citalopram (CIT) in patients with severe depression (Montgomery-Asberg depression rating scale [MADRS] score ≥30)*

<table>
<thead>
<tr>
<th>Description</th>
<th>Week</th>
<th>Moore (9)* (ESC n = 138; CIT n = 142)</th>
<th>Llorca (13), Colonna (8) (pooled ESC n = 249; CIT n = 260)*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECS-CIT MADRS reduction</strong></td>
<td>4</td>
<td>Mean difference (95% CI) = -0.7 (-2.52 to 1.12)</td>
<td>Data on patients with MADRS≥30 selected from other RCTs with MADRS≥22 at inclusion</td>
</tr>
<tr>
<td>Difference &lt;0 favors ESC</td>
<td></td>
<td>Standardized (95% CI) = -0.090 (-0.324 to 0.145)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Mean difference (95% CI) = -2.1 (-4.21 to 0.01)</td>
<td>Pooled random WMD (95% CI) = -1.901 (-5.407 to 1.605)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standardized (95% CI) = -0.233 (-0.468 to 0.002)</td>
<td>Pooled random SMD (95% CI) = -0.172 (-0.490 to 0.146)</td>
</tr>
<tr>
<td><strong>ESC-CIT CGI-S reduction</strong></td>
<td>8</td>
<td>Mean difference (95% CI) = -0.120 (-0.455 to 0.215)</td>
<td>Pooled random SMD (95% CI) = -0.220 (-0.640 to 0.199)</td>
</tr>
<tr>
<td>Difference &lt;0 favors ESC</td>
<td></td>
<td>Standardized (95% CI) = -0.084 (-0.318 to 0.151)</td>
<td></td>
</tr>
<tr>
<td><strong>ESC vs CIT response (MADRS)</strong></td>
<td>8</td>
<td>76.1% vs 61.3%</td>
<td>Pooled random proportions 54.6% vs 45.7%</td>
</tr>
<tr>
<td>RR&gt;1 favors ESC</td>
<td></td>
<td>RR (95% CI) = 1.24 (1.06 to 1.47)</td>
<td>Pooled random RR (95% CI) = 1.184 (0.861 to 1.627)</td>
</tr>
<tr>
<td>NNT = 7 benefit (4 benefit to 25 benefit)</td>
<td></td>
<td>Pooled random NNT = 12 benefit (4 benefit to 14 harm)</td>
<td></td>
</tr>
</tbody>
</table>

*Abbreviations: RCT – randomized controlled trial; WMD – weighted mean difference; SMD – standardized mean difference; CGI-S – Clinical Global Impression-severity (35); RR – relative risk; NNT – number needed to treat; CI – confidence interval.

†One of the trials included in the main efficacy analysis.

‡Patients included in these reports are selected from RCTs based on MADRS (33). None of the RCTs employed stratified randomization in respect to disease severity. Llorca (13) – post-hoc analysis of selected pooled data from Burke (6), Lepola (7), and SCT-MD-02 (11) (ESC n = 169, CIT n = 171); Colonna (8) – a subgroup of patients from an RCT (ESC n = 80, CIT n = 89).
DISCUSSION

Why look into escitalopram vs citalopram clinical trial data again?

Depression is common, seriously disabling, chronic in nature and expensive disease – it is one of the leading causes of loss of disability-adjusted life-years world-wide with total costs amounting to hundreds of millions in any country or currency (1). Development of newer antidepressants has enabled wider use of pharmacological treatments for depression. This valuable contribution has, however, turned depression into a premium marketing playground; most of the newer antidepressants are among top selling drugs worldwide with individual sales in billions (30). At the same time, there is growing evidence that benefits of antidepressants have been considerably overestimated and that decisions on “who should be treated” and “which drug should be used” have been largely influenced by marketing efforts and less by evidence (28–30). In this context, the “story of escitalopram and citalopram” is a specific one since both were developed and marketed by the same company. While there is a sound molecular rationale (2) to explain why should one expect more from the “pure active enantiomer” (escitalopram) than from “active entantiomer in a racemic mixture” (citalopram) on equimolar basis, translation of this “marker-based reasoning” into clinical evidence is inevitably marketing-driven. Escitalopram came around at the time of expiration of exclusivity of citalopram and in order to keep the market share the manufacturer actually had to offer “something new and better.” Individually, trials comparing escitalopram and citalopram share common characteristics of clinical trials of antidepressants in general (30) – they are mostly short and small – and their results are inconclusive regarding two important questions: a) What is the actual size of the between-drug differences? and b) What is their practical/clinical relevance? In an attempt to contribute to answering these questions, the present analysis had specific aims. First, it compared equimolar (S-[+] enantiomer) doses. This decision was based on the following: a) the formal dosing recommendations for the two drugs reflect the concept of “proportionality” (3,4); b) although in practice the two drugs are likely used after other different modalities, there is no RCT that actually aimed to answer the question whether the “free (intuitive, unrestrained) use” of one drug provided a possibility of a better general control of depression than the other drug. Under such circumstances, inclusion of treatment arms with “non-equimolar” doses in between-drug comparisons could be a source of noise compromising the estimation of size of the differences. Second, it included CGI-S as an efficacy outcome. Although CGI-S is not appropriate to serve as a single instrument in trials evaluating potential antidepressant compounds, it is sensitive enough as to capture the effects of antidepressants and provides robust information on changes in clinical sta-
tus of depressed patients that is useful for evaluations in daily practice (36,42,43). In this respect, it is a valuable aid in an attempt to estimate clinical relevance of differences between two established antidepressants. Third, it assesses differences at several time points. Although this could be a source of a multiplicity problem (31), it could also allow for a more reliable assessment of treatment differences than comparisons based on a single-point measurement of an outcome susceptible to fluctuation.

Size and practical relevance of efficacy differences between escitalopram and citalopram

Assessing clinical relevance is a largely arbitrary effort particularly in depression (30), but the criteria used in the present assessment referring to MADRS reduction, NNT, and effect size are generally accepted as reasonable. The only arbitrary criterion was that referring to the difference in CGI-S reduction – if ≤0.32 scale points (ie, less than 1/3 of a point) then it is irrelevant. With CGI-S being an ordinal scale, it seems reasonable to consider that in order to fall into the next upper or lower level (category), the "severity" needs to change to the extent that reaches at least "half the way" toward this next step (ie, "half a point"). Hence, excursions (from the actual level) that are less than 1/3 of the "way towards the next level" are unlikely to result in a change in CGI-S grading and could be considered not clinically relevant.

In line with the molecular background, escitalopram provided greater MADRS reduction, higher proportion of responders and, less so, a greater reduction in CGI-S. By their sizes, however, all the differences could be judged as being below clinical/practical relevance. Data are most abundant for MADRS reduction. At all time points (1, 4, 6, 8, and 24 weeks), 95% CI around effect sizes were within the range of "small." Similarly, 95% CI around mean differences were consistently entirely (or nearly so) below the limit of a "minimally important difference" (2 points). Only at week 6 was the pooled estimate "shifted" somewhat toward the limit of "relevance." It should be noted, however, that in one of the trials (8) that was conceived as a non-inferiority trial (for escitalopram vs citalopram), the acceptance limit for difference in MADRS was set at 3 points. By this criterion, and with a "reverse logic," the present data suggest that citalopram is non-inferior to escitalopram regarding MADRS reduction – over the entire short-to-medium-term treatment. Regarding the response rates and CGI-S reduction, only at week 8 were the comparisons based on a reasonable amount of data. Escitalopram is apparently associated with a higher "risk" of being a responder by week 8 of treatment. Although inconsistency in meta-analysis of response rates was mild, this conclusion is largely influenced by the trial that included only severely depressed patients (9). However, even this estimate does not support a conclusion of practical relevance since NNTB was 14 with 95% CI extending from 7 to 111. If the estimate is accurate, additional studies in, eg, a prospective cumulative meta-analysis, would improve its precision (shorten the confidence interval), but point-estimate is not likely to fall below 10. The findings of a recent network meta-analysis (21) (ie, a direct-indirect comparison based on a larger number of trials) support such a conclusion – it yielded an odds ratio of "response" that corresponds to an NNT of 25 (assuming the citalopram response rate of 50%) with CI extending from benefit to harm.

The differences between two treatments were lowest regarding reduction of CGI-S. Confidence intervals around pooled WMD and effect size were not only below 0.32, but were below 0.25 (less than 1/4 of a point or far lower than the limit of a "small" effect size). The escitalopram vs citalopram differences in all 3 efficacy measures appeared greatest at week 6. They were all largely influenced by one same 6-week trial conducted in Russia (Yevtushenko et al [10]). This particular trial differed from all the others providing 6-week or 8-week data in that MADRS reduction, response rates, and CGI-S reduction were far greater for both drugs, and in that the differences between the two drugs were far larger. The specificity of the trial was that it included only patients 25-45 years of age with MADRS≥25. In that respect, the results could actually be more informative about the sample than about the general between-treatment relationship.

The basis for evaluation of escitalopram and citalopram in severely depressed patients (typically, MADRS≥30) is very modest – there is only one RCT specifically addressing the issue (9). It showed a greater "risk" of being a responder at week 8 associated with escitalopram, but did not unequivocally support a conclusion of relevant superiority. NNTB was 7 (95% CI, 4 to 25), thus being "close to relevant;" but there was practically no difference between treatments in CGI-S reduction, practically no difference in MADRS reduction at weeks 1 and 4, whereas the difference at week 8 (and particularly the standardized difference) was largely below the limits of relevance. Other data on "severe patients" (extensively exploited in various reviews) are actually only supportive: one of the trials (8) reported on a subgroup of such patients and an additional report (13) was a
post-hoc analysis of patients selected from 3 other RCTs (all trials without stratified randomization). The present pooled estimates based on this data were in favor of escitalopram, but did not, by size, support a conclusion of “relevance.” Another analysis of these same data (12) demonstrated a significant trend of increasing between-treatment differences in patients with higher MADRS. For example, for a small subgroup of patients (90 in total) with MADRS≥35, escitalopram vs citalopram difference in MADRS reduction at week 8 was estimated to be 6 points. However, the confidence interval extended from around 0.1 to 13. Overall, it appears plausible to conclude that in more severely ill patients the differences in efficacy between escitalopram and citalopram could be expected to be larger – still, it does not mean that they would be clinically/practically relevant.

Comparative safety and discontinuations due to AE and inefficacy

The considered trials are by size and duration by far too small of a basis for “final” evaluation of safety of the two drugs. However, they are useful for a direct safety profile comparison “under identical conditions.” A detailed analysis by the Cochrane group (22) showed no relevant difference between the two drugs in respect to total incidence of AEs or incidence of any individual AE. For antidepressants, discontinuation of treatment is a major factor of success or failure since the disease typically requires sustained treatment. In a drug comparison setting, discontinuations due to AE and/or inefficacy are particularly interesting as they are directly “linked” to the drug and could be accurately recorded in formal double-blind RCTs. The Cochrane meta-analysis (22) reported on these outcomes (with no significant or relevant difference between the drugs), but pooled together all trial data irrespective of their duration, included also “non-equimolar” doses and used “all randomized” patients as a basis. The present overview separates discontinuations during the short-term (up to 8 weeks) and medium-term (week 9 to 24) treatment, stays with equimolar doses (which were also compared for efficacy), and uses “safety” data sets as the basis. Under these conditions, there was no relevant difference between the treatments regarding short-term discontinuations due to AE or inefficacy (cumulatively) or due to AEs. Discontinuations due to inefficacy were almost twice more frequent with citalopram, but the numbers were very low (0.4% and 0.9%) and inconclusive. Data for week 9 to 24 are actually limited to around 300 patients that completed the first 8 weeks in the single 24-week trial (8) and are too scarce for any reasonable conclusion.

Clinical relevance is not only about effect size

We seem to have built (or have allowed it to be built) a “culture of antidepressant clinical trials” which, due to a number of reasons (overall trial design and analysis issues, patient selection, (un)publishing policies, data “furnishing,” regulatory (non)submissions, industry influence), produces results that do not reflect the real physical world that these drugs are intended for (30,44). First, effects sizes are inflated suggesting that there is more benefit in these drugs than there really is. Second, generalizability of these trials is, at best, very modest – according to a “typical” RCT of an antidepressant (and this applies to the trials addressed in this overview, as well), the drug would actually be intended only for generally younger, moderately to severely ill MDD patients (as assessed on a depression measuring scale), with moderately long disease history, who, in general, had not received previous treatment and are free of other psychopathology or comorbidity (30,44). Comparing antidepressants based on such trials with the aim of selecting the “right one” for the general population is inherently problematic (44). A truly meaningful evaluation of antidepressants in general and in comparisons would require pragmatic trials that are compliant with the nature of the disease and characteristics of the targeted population (30,44). Hence, in respect to the “general population,” clinical relevance of escitalopram vs citalopram differences in treatment of depression is limited not only by the small effect sizes but also by the general characteristics of the existing trials.

In conclusion, existing RCTs do not support the claim about clinically relevant superiority of escitalopram over citalopram (on equimolar basis) in short-term (4-8 weeks) or medium-term (up to 24 weeks) treatment of MDD. Data addressing specifically only “severe” patients (MADRS≥30) are scarce and currently inconclusive. They do indicate a possibility of consistent and larger differences in favor of escitalopram, but this is yet to be formally demonstrated.

Acknowledgment

This study received no funding. I am a member of the Committee for Human Medicines at the Agency for Medicines and Medical Devices of the Republic of Croatia. Over the past 10 years I have provided scientific consultation to and have received honoraria from pharmaceutical industry (Baxter Healthcare International, Fumedica GmbH, APPH GmbH, Pentapharma Ltd, Sandoz-Lek, Pliva, Belupo, JGL, and Janssen-Cilag Croatia). I have no other conflict of interest to declare.
References


Trkulja: Escitalopram vs Citalopram for Depression


