Efficacy of Second Generation Antidepressants in Late-Life Depression: A Meta-Analysis of the Evidence

J. Craig Nelson, M.D., Kevin Delucchi, Ph.D., Lon S. Schneider, M.D.

Objective: Second-generation antidepressants are commonly used to treat major depression in late-life. This systematic review and meta-analysis was undertaken to assess the evidence for efficacy of second-generation antidepressants in late-life major depression. Methods: The Cochrane Library (2006 [3]), MEDLINE (1966 to August 2006), and meeting presentations were searched for trials of second-generation antidepressants (nontricyclics) marketed in the United States. Published and unpublished placebo-controlled randomized clinical trials in outpatients 60 years and older, with nonpsychotic, unipolar major depression were selected. Clinical characteristics and outcomes were extracted. Outcomes were expressed as odds ratios (OR), risk differences, and weighted mean differences. Results: Ten unique trials (four unpublished) with 13 contrasts met selection criteria. Trials were 6–12 weeks duration, and included 2,377 patients who received active drug and 1,788 received placebo. The ORs by meta-analysis for response and remission were 1.40 (95% confidence interval [CI] 1.24–1.57, z = 5.45, N = 13, p < 0.001) and 1.27 (CI 1.12–1.44, z = 3.67, N = 13, p < 0.001), respectively, with significant heterogeneity for response and remission among the trials. Mean pooled response rates for antidepressant and placebo were 44.4% and 34.7%, respectively. The OR for response was significantly higher in the 10–12 week trials (OR = 1.73, CI 1.42–2.09, z = 5.51, N = 5, p < 0.001) than the 6–8 week trials (OR = 1.22, CI 1.05–1.42, z = 2.60, N = 8, p = 0.01). ORs for discontinuation for any reason and for adverse events were significantly higher with drugs than with placebo. Conclusions: Antidepressants are more effective than placebo in elderly depressed subjects although effects are modest and vary. Identification of the characteristics of responders and nonresponders will be crucial to improving treatment outcomes. (Am J Geriatr Psychiatry 2008; 16:558–567)

Key Words: Late-life depression, antidepressants, meta-analysis
Selective serotonin reuptake inhibitors (SSRIs) and other second-generation antidepressant medications, including bupropion, mirtazapine, venlafaxine, and duloxetine are the most commonly prescribed pharmacological treatments for depressive disorders in late-life and are recommended as first line treatments in clinical practice guidelines. Although previous systematic reviews of antidepressant drugs found evidence for efficacy in older adults with depression, reviews or meta-analyses before 2003 focused principally on the tricyclic antidepressants (TCAs) because until that year only one large placebo-controlled trial of a nontricyclic antidepressant, marketed in the US, in outpatients 60 years of age or older with major depressive disorder had been published. Because the first few trials of second-generation agents reported no advantage relative to placebo or small drug-placebo differences, the efficacy of these agents in older depressed adults was questioned.

A recent review assessed psychotherapy and various medications across a range of depression-related diagnoses in middle- and older-aged adults. The current review and analysis differs from that study. We focused on the evidence for efficacy of second-generation antidepressants for the primary indication—Major Depression. We restricted our analysis to trials in subjects 60 years and older and limited the search to antidepressants approved for use in the United States. This analysis is timely because nine placebo-controlled studies of second-generation drugs have been reported or presented since 2003 and no previous review or meta-analysis has included more than four of these recent trials.

METHODS

The Cochrane Controlled Trials Register (2006, Issue 3) was searched using terms previously employed by the Cochrane group in a similar search (elder, geriatr, senil, older, old age, late-life, aged, 80-and-over, and depress). These terms were supplemented with the terms antidepressants, fluoxetine, sertraline, paroxetine, citalopram, escitalopram, venlafaxine, duloxetine, mirtazapine, bupropion, nefazodone, and trazodone (second-generation antidepressants marketed in the United States). MEDLINE (1966 to August 2006) was also searched. Proceedings from geriatric psychiatric and psychiatric professional society meetings since 2000, and previous reviews were hand-searched. Pharmaceutical manufacturers were queried and information was requested as needed.

Trials Selection

Trials were included if they met the following criteria: acute phase, parallel group, double-blinded, placebo-controlled with random assignment to an orally administered second generation antidepressants (nontricyclics) marketed in the United States; patients had nonpsychotic, unipolar Major Depression not associated with a specific medical disorder, e.g., poststroke depression or dementia; patients were community dwelling, and aged 60 years or older; numbers of patients randomized, outcomes, and dropouts were obtainable. Trials did not need to be published or peer reviewed and could be reported in manuscripts, technical trials reports, or posters. (Some sources presented incomplete information regarding remission rates, mean change scores with the standard deviations (SD), or adverse event discontinuation rates and this information was obtained though other data presentations or from sponsors). Two authors identified and agreed upon studies meeting these criteria. Study quality was assessed with Jadad scores.

Data Extraction

Information extracted included study design, patient selection criteria, medication doses, location, trial durations, age, sex, baseline depression rating scores, numbers randomized, clinical outcomes on rating scales, and dropouts occurring during the double-blind trial period.

Clinical outcomes included response, remission, and change scores on the Hamilton Depression Rating Scale (HAMD) or Montgomery Asberg Depression Rating Scale (MADRS) and were assessed in the intent-to-treat (ITT) samples using the last observation carried forward in patients with at least one posttreatment rating. Response was defined as ≥50% improvement from baseline on the HAMD or the MADRS. Remission was defined by the individual study. In flexible dose trials that compared two formulations of the same medication, we combined the drug groups to make one contrast with the placebo group to minimize multiple comparisons with the same placebo group. For fixed-dose trials, comparing two doses of a medication, we
compared each dosage group to placebo. In trials with an active comparator we compared each drug within the trial to the placebo group. Data were abstracted by one investigator (JCN) and checked by another investigator (LSS). Any discrepant data were rereviewed by the investigators to ensure that accurate data were obtained.

Necessary information not included in the publication or presentation was requested from the investigator or sponsor of the study. If SD of the change scores were not available, they were estimated from standard errors and sample sizes or imputed using the largest SD reported in other trials.

Statistical Analysis

The number of responders, remitters, dropouts, and the number randomized into each drug and placebo group for each trial were statistically combined using the Peto fixed-effects model. Effects were expressed as odds ratios (ORs) and absolute risk differences (RDs) with their 95% confidence intervals (CIs), test of significance (Wald z), number (N) of contrasts, and p values. The OR is the odds of an event (e.g., response) occurring in one group divided by the odds of the event in the other group (e.g., the placebo group). The overall ORs for the meta-analysis is the mean of the ORs computed for each contrast weighted for sample size and the event rate. The RD for the meta-analysis is the mean of the differences between the risk of the event (response) in one group minus that in the other group weighted as above. Mean change scores on the HAMD for each drug and placebo comparison were combined using an inverse variance fixed-effects method. Effects were expressed as weighted mean differences (WMDs) with the 95% CIs, z score, N, and p values. The WMD is the difference between the mean change scores of the treatment and control groups weighted by the sample size. Effects were calculated for each drug-placebo contrast, for each drug class (SSRI, serotonin norepinephrine reuptake inhibitor, e.g., venlafaxine; dopamine norepinephrine reuptake inhibitor, e.g., bupropion) separately, and as meta-analytic summaries for all drugs combined. A funnel plot in which the standard error (SE) of the log OR against the log OR was used to evaluate potential retrieval bias.

χ^2 tests and the I^2 statistic derived from the χ^2 values were used to test heterogeneity among the contrasts. I^2 approximates the proportion of total variation in the effect size estimates that is due to heterogeneity rather than sampling error. An alpha error p < 0.20 and I^2 of at least 50% were taken as indicators of heterogeneity of outcomes. In the text, the χ^2 test and I^2 statistics for heterogeneity follow the 95% CI, z score, N, and p value for the ORs unless these values are provided in a figure. For outcomes that were heterogeneous we examined the effect of individual removal of extreme outcomes on the I^2 statistic.

We compared the following subgroups as sensitivity analyses: potential differences between SSRIs and other drugs and between studies of 6–8 week versus 10–12 week duration. Differences between two or more subgroups were investigated by subtracting the sum of the heterogeneity χ^2 statistics of the subgroups from the overall χ^2 statistic and comparing the result with a χ^2 distribution with a df of one less than the number of subgroups. Review Manager version 4.2 (The Cochrane Collaboration, Oxford, England) was used for statistical calculations.

Role of Funding Source

No external funding was received for the study design, trial search, data analysis, interpretation of the data, writing of the paper, or the decision to submit for publication.

RESULTS

Search Flow (Results)

The search strategy yielded 192 Cochrane Controlled Trials Register citations and 136 citations from MEDLINE (Fig. 1). The Cochrane search identified 11 placebo-controlled studies and the MEDLINE search 25 placebo-controlled studies. Of these, five trials from the Cochrane search and six trials from the MEDLINE search met inclusion criteria. (The MEDLINE search included all of the Cochrane trials plus one recently published report). A search of abstracts, posters, and slide presentations from medical conferences revealed four other placebo-controlled studies.

Trials Description

Ten trials with 13 contrasts (fluoxetine 3, escitalopram 2, sertraline 1, paroxetine 3, citalopram 1,
venlafaxine 1, duloxetine 1, and bupropion 1) met the selection criteria (Table 1).9–18 All trials included patients with nonpsychotic, unipolar Major Depression. They were 6–12 weeks in duration, included 178–752 patients per trial; mean age ranged from 68 to 80 years; mean proportion women from 46% to 76%. None of these trials included a specific psychotherapy comparison. All reported ITT sample size and outcomes or that data were obtained. All used the Mini-Mental State Examination to exclude patients with cognitive impairment, but the entry criterion varied from ≥25 to ≥18. Eight trials used the HAMD and two used the MADRS to assess outcome. Mean baseline HAMD scores ranged from 19 to 24 indicating moderate severity of depressive symptoms on average. All trials were sponsored by the manufacturer of the antidepressant, and had Jadad scores of 4 or 5 indicating that the trials were of good to excellent methodological quality.

Clinical Outcomes

Response rates in the trials ranged from 35% to 69% for the antidepressant and from 19% to 47% for placebo. Six of the 10 trials reported statistically significant advantages for drug compared with placebo on response. The overall OR by meta-analysis for response with medications was 1.40 (95% CI 1.24–1.57, z = 5.45, N = 13, p < 0.001) (Fig. 2). RD by meta-analysis was 0.08 (95% CI, 0.05–0.11, z = 5.51, N = 13, p < 0.001). The pooled response rate for drug was 44.4% versus 34.7% for placebo.

Remission rates for the antidepressant ranged from 21% to 44% and from 13% to 42% for placebo. The overall OR by meta-analysis for remission with medications was also significantly higher, (OR = 1.27, CI 1.12–1.44, z = 3.67, N = 13, p < 0.001) (Fig. 3). RD by meta-analysis was 0.05 (95% CI, 0.02–0.08, z = 3.71, N = 13, p < 0.001). The pooled remission rate was 32.6% for drug versus 26.5% for placebo.

Change scores in the eight trials (and 10 contrasts) that reported HAMD scores were greater for medication, WMD = 1.40 (95% CI 0.89–1.90, z = 5.39, N = 10, p < 0.001; heterogeneity $\chi^2 = 19.27, df = 9, p = 0.02, I^2 = 53.3\%$).

Discontinuation rates varied from 17% to 36% for the drug groups and 11%–30% for placebo. The OR by meta-analysis indicated increased odds for discontinuation on medication (OR = 1.22, 95% CI 1.06–1.40, z = 2.80, N = 13, p = 0.005; $\chi^2 = 23.15, df = 12, p = 0.03, I^2 = 48.2\%$). Pooled mean discontinuation rates were 24% and 20% for drug and placebo, and RD by meta-analysis was 0.03 (95% CI 0.01–0.06, p = 0.005). Discontinuation rates due to adverse effects varied from 8% to 27% for medication and 1%–11% for placebo. The OR for discontinuations due to adverse events on medication was 1.84 (95% CI 1.51–2.24, z = 6.02, N = 13, p < 0.001; $\chi^2 = 30.82, df = 12, p = 0.002, I^2 = 61.1\%$). Pooled mean adverse event dropout rates were 12% and 7% for drug and placebo, and RD by meta-analysis was 0.05 (95% CI 0.03–0.07, z = 6.10, N = 13, p < 0.001).

In eight trials (10 contrasts) that included an SSRI in one of the treatment arms, likelihood of response was higher for SSRIs than placebo (OR = 1.36, 95% CI 1.19–1.56, z = 4.55, N = 10, p < 0.001; $\chi^2 = 25.45, df = 9, p = 0.003, I^2 = 64.6\%$) (Fig. 2) and the ORs for the SSRIs and the non-SSRI agents seemed similar (had overlapping CIs). The WMD on the HAMD for the
<table>
<thead>
<tr>
<th>Study Authors, Year</th>
<th>Patient Selection Criteria</th>
<th>Age Mean (SD)</th>
<th>Sex % Female</th>
<th>Duration Weeks</th>
<th>Baseline Severity</th>
<th>Drug</th>
<th>N* ITT</th>
<th>Dropout N (%)</th>
<th>N* ITT/LOCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tollefson et al 1995⁹</td>
<td>DSM III-R MDD; ≥ 4 weeks; ≥60 years; HAMD17 ≥16; MMSE ≥ 25</td>
<td>67.8 (5.7)</td>
<td>54.5</td>
<td>6</td>
<td>HAMD17</td>
<td>22.2 (3.9)</td>
<td>Fluoxetine</td>
<td>335</td>
<td>72 (25.5)</td>
</tr>
<tr>
<td>Schneider et al 2003¹²</td>
<td>DSM IV MDD; ≥4 weeks; ≥60 years; HAMD17 ≥18; MMSE ≥ 24</td>
<td>69.8 (6.6)</td>
<td>56.0</td>
<td>8</td>
<td>HAMD17</td>
<td>21.4 (2.7)</td>
<td>Sertraline</td>
<td>371</td>
<td>87 (23.4)</td>
</tr>
<tr>
<td>Rapaport et al 2003¹³</td>
<td>DSM IV MDD; ≥60 years; HAMD17 ≥18; MMSE ≥ 25</td>
<td>70.0 (6.0)</td>
<td>56.0</td>
<td>12</td>
<td>HAMD17</td>
<td>22.1 (3.4)</td>
<td>Paroxetine CR</td>
<td>104</td>
<td>23 (22.1)</td>
</tr>
<tr>
<td>Roose et al 2004¹⁰</td>
<td>DSM IV MDD; ≥ 4 weeks; ≥74 years; HAMD24 ≥ 20; MMSE ≥ 18</td>
<td>79.6 (4.4)</td>
<td>58.1</td>
<td>8</td>
<td>HAMD24</td>
<td>22.1 (3.0)</td>
<td>Citalopram</td>
<td>109</td>
<td>25 (22.9)</td>
</tr>
<tr>
<td>Kasper et al 2005¹¹</td>
<td>DSM IV MDD; ≥ 4 weeks; ≥65 years; MADRS ≥ 22; ≥40; MMSE ≥ 22</td>
<td>75 (7)</td>
<td>76.0</td>
<td>8</td>
<td>MADRS</td>
<td>28.2 (3.8)</td>
<td>Escitalopram</td>
<td>173</td>
<td>29 (16.8)</td>
</tr>
<tr>
<td>Raskin et al 2005 (poster)¹⁴</td>
<td>DSM IV MDD; recurrent; ≥65 years; HAMD17 ≥18; MMSE ≥ 20</td>
<td>72.8 (5.6)</td>
<td>59.5</td>
<td>8</td>
<td>HAMD17</td>
<td>18.8 (4.8)</td>
<td>Placebo</td>
<td>180</td>
<td>20 (11.1)</td>
</tr>
<tr>
<td>Schatzberg et al 2006¹⁵</td>
<td>DSM IV MDD; ≥4 wks; ≥65 years; HAMD21 ≥ 20; MMSE ≥ 18</td>
<td>71 (5.0)</td>
<td>49.1</td>
<td>8</td>
<td>HAMD21</td>
<td>24 (3)</td>
<td>Venlafaxine IR</td>
<td>104</td>
<td>37 (36)</td>
</tr>
<tr>
<td>Bose et al 2006 (poster)¹⁶</td>
<td>DSM IV MDD; ≥ 60 years; MADRS ≥ 22; MMSE ≥ 24</td>
<td>68.3 (6.9)</td>
<td>59.4</td>
<td>12</td>
<td>HAMD17</td>
<td>24.3 (4.6)</td>
<td>Placebo</td>
<td>134</td>
<td>25 (18.7)</td>
</tr>
<tr>
<td>Rousseau et al 2006 (poster)¹⁷</td>
<td>DSM IV MDD; ≥65 years; HAMD17 ≥ 18; MMSE ≥ 25</td>
<td>71.1</td>
<td>72%</td>
<td>10</td>
<td>MADRS</td>
<td>29.5 (4.9)</td>
<td>Bupropion XL</td>
<td>211</td>
<td>49 (23)</td>
</tr>
<tr>
<td>Pitts et al 2005 (poster)¹⁸</td>
<td>DSM IV MDD; ≥60 years; HAMD17 ≥ 18; MMSE ≥ 25</td>
<td>67 (6.5)</td>
<td>61%</td>
<td>10</td>
<td>HAMD17</td>
<td>22.6 (3.6)</td>
<td>Paroxetine CR 12.5 mg/d</td>
<td>164</td>
<td>39 (23.8)</td>
</tr>
</tbody>
</table>

* N ITT includes all subjects randomized and was used in calculation of discontinuation rates. N ITT/LOCF includes all patents randomized that completed at least one post-treatment assessment.
SSRI trials was 1.37 (95% CI 0.84–1.90, z = 5.06, N = 8, p < 0.001; χ² = 16.65, df = 7, p = 0.02, I² = 58.0%). Six of the trials were of 6–8 weeks duration whereas four were 10–12 weeks. Response rates were higher in the longer trials. The OR for response by meta-analysis for the 10–12 week trials was 1.73 (95% CI 1.42–2.09, z = 5.51, N = 5, p < 0.001) whereas that for the 6–8 week studies was 1.22 (95% CI 1.05–1.42, z = 2.60, N = 8, p = 0.01) (Fig. 4) and the difference between the ORs was significant (χ² = 7.42, df = 1, p < 0.01). Pooled mean response rates for drug and placebo groups respectively were 55% versus 41% in the longer trials and were 38% versus 31% in the shorter trials, with RDs, by meta-analysis of 0.14 (95% CI 0.09–0.18, z = 5.57, N = 5, p < 0.001) and 0.05 (95% CI 0.01–0.08, z = 2.63, N = 8, p = 0.009) for the longer and shorter trials, respectively.

Removal of the trial with the lowest OR, a comparison of escitalopram, fluoxetine, and placebo,11 substantially decreased the I² statistic, an estimate of the proportion of the variance due to heterogeneity, from 62.3% to 32.8% although the χ² test remained significant (χ² = 14.88, df = 10, p = 0.14) using the a priori alpha level of p < 0.20. Removal of the trial with the highest OR14 had little effect on I² (change from 62.3% to 59.9%). A funnel plot of the SE of the log OR against the log OR revealed a symmetrical distribution about the mean indicating no evidence of publication bias.

**DISCUSSION**

Second-generation antidepressants are more effective than placebo during acute treatment of adults 60
Antidepressants in Late-Life Depression

years and older with Major Depression in terms of response and remission defined by depression rating scale scores, but the magnitude of this effect is small and variable. The numbers needed to treat calculated from the RDs by meta-analysis were 13 for response and 20 for remission, implying that for every 100 patients treated 8 would show a response and 5 a remission in excess of placebo treatment. The benefits of drug treatment, however, should be weighed against the rates of adverse events requiring discontinuation. For every two patients who responded to drug treatment, one discontinued prematurely because of adverse effects.

Short trials of 6–8 weeks seem to underestimate outcomes achieved in longer trials. The mean response rates in the 6- and 8-week trials ranged from 35% to 46.5%, the OR for response was 1.22, and the RD was 5%. In the 10- to 12-week trials, response rates ranged from 45.7% to 68.9%, the OR was 1.73, with a RD of 14%. This observation, however, is based on only four trials (five contrasts) of 10–12 weeks duration and four of the five contrasts in the longer trials were with drugs not tested in the shorter trials.

These small drug-placebo differences might raise questions about the relative efficacy of the second-generation drugs compared with the older TCAs. A Cochrane meta-analysis of TCA and SSRI comparison trials in older patients found no advantage of TCAs compared with SSRIs in five trials with approximately 400 patients receiving each drug class, although these trials were not placebo-controlled.25 We note, however, that the number of comparisons is again small.

In the studies reviewed, response rates with placebo varied from 19% to 47%. Randomization to
placebo is employed in depression trials to control for spontaneous improvement, treatment expectations, and nonspecific factors that might affect outcomes. Patients in clinical trials are seen for frequent visits and receive education, attention, reassurance, and clinical monitoring—interventions which have therapeutic effects. Clinicians should recognize the important contribution of these therapeutic interventions to overall outcome.

Because evidence for heterogeneity of outcomes was observed in the meta-analysis, the findings should be interpreted by considering not only the overall result, but also the individual study results. Results of one trial\textsuperscript{11} contributed about half the proportion of variance due to heterogeneity. This 8-week trial with escitalopram and fluoxetine was different from the others in that it was performed at a very large number of sites (76 sites) and had the highest placebo response rate, 47%.

Heterogeneity of outcomes could be due to many factors including differences in patient selection, diagnoses, study design, ascertainment of the outcomes, differences among the drugs, and often the sources of heterogeneity cannot be determined. The 10 trials assessed eight antidepressants. Although the ORs for individual drugs varied (Fig. 2), no single medication or class was superior to another as evidenced by the overlapping ORs. But any interpretation of equivalent efficacy would be based on sparse data.

Patient factors such as age, past history, depression severity, and cognitive function varied in the samples and may influence outcome. As examples, the citalopram trial\textsuperscript{10} included the oldest patients and found no drug-placebo difference in response, but in a post-hoc analysis of that trial, greater drug-placebo differences were found among the most severe patients. One trial with a large drug-placebo difference restricted selection to patients with recurrent depression,\textsuperscript{14} and 3 of the 4 trials that allowed patients with significant cognitive impairment (i.e., Mini-Mental State Exam scores below 24) had nonsignificant outcomes.\textsuperscript{10,11,15}

![FIGURE 4. ITT/LOCF Response Rates by Individual Trial, by Study Duration, and Overall Compared With Placebo](image-url)

<table>
<thead>
<tr>
<th>Study or sub-category</th>
<th>Treatment</th>
<th>Control</th>
<th>Peto OR 95% CI</th>
<th>Weight</th>
<th>Peto OR 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 6-8 week studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kasper escitalopram</td>
<td>79/170</td>
<td>85/180</td>
<td>8.17 (0.64, 1.48)</td>
<td>6.17</td>
<td>6.17 (0.64, 1.48)</td>
</tr>
<tr>
<td>Kasper fluoxetine</td>
<td>61/164</td>
<td>85/180</td>
<td>7.87 (0.43, 1.02)</td>
<td>6.87</td>
<td>6.87 (0.43, 1.02)</td>
</tr>
<tr>
<td>Raskin duloxetine</td>
<td>76/201</td>
<td>39/102</td>
<td>6.44 (2.39, 1.59)</td>
<td>6.44</td>
<td>6.44 (2.39, 1.59)</td>
</tr>
<tr>
<td>Roose citalopram</td>
<td>34/84</td>
<td>34/90</td>
<td>3.89 (0.61, 2.06)</td>
<td>3.89</td>
<td>3.89 (0.61, 2.06)</td>
</tr>
<tr>
<td>Schatzberg fluoxetine</td>
<td>39/99</td>
<td>40/96</td>
<td>4.62 (0.51, 1.61)</td>
<td>4.62</td>
<td>4.62 (0.51, 1.61)</td>
</tr>
<tr>
<td>Schatzberg venlafaxine</td>
<td>37/93</td>
<td>40/96</td>
<td>4.72 (0.32, 1.65)</td>
<td>4.72</td>
<td>4.72 (0.32, 1.65)</td>
</tr>
<tr>
<td>Schneider sertraline</td>
<td>126/360</td>
<td>96/368</td>
<td>14.45 (1.33, 2.09)</td>
<td>14.45</td>
<td>14.45 (1.33, 2.09)</td>
</tr>
<tr>
<td>Tollefson fluoxetine</td>
<td>117/325</td>
<td>89/329</td>
<td>13.22 (1.09, 2.10)</td>
<td>13.22</td>
<td>13.22 (1.09, 2.10)</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>1496</td>
<td>1441</td>
<td>61.74 (1.05, 1.43)</td>
<td>61.74</td>
<td>61.74 (1.05, 1.43)</td>
</tr>
<tr>
<td>Total events: 598 (Treatment), 488 (Control)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for heterogeneity: $\chi^2 = 20.96, df = 7 (P = 0.004), F = 66.6%$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 2.60 (P = 0.009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>02 10-12 weeks studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bose escitalopram</td>
<td>59/129</td>
<td>51/134</td>
<td>6.01 (0.84, 2.23)</td>
<td>6.01</td>
<td>6.01 (0.84, 2.23)</td>
</tr>
<tr>
<td>Pitts paroxetine 12.5</td>
<td>85/163</td>
<td>72/179</td>
<td>7.96 (1.06, 2.57)</td>
<td>7.96</td>
<td>7.96 (1.06, 2.57)</td>
</tr>
<tr>
<td>Pitts paroxetine 25</td>
<td>100/173</td>
<td>72/179</td>
<td>8.25 (1.33, 3.06)</td>
<td>8.25</td>
<td>8.25 (1.33, 3.06)</td>
</tr>
<tr>
<td>Rapaport paroxetine</td>
<td>135/206</td>
<td>47/107</td>
<td>6.46 (1.45, 3.72)</td>
<td>6.46</td>
<td>6.46 (1.45, 3.72)</td>
</tr>
<tr>
<td>Rousseau bupropion</td>
<td>110/207</td>
<td>87/203</td>
<td>9.60 (1.62, 2.22)</td>
<td>9.60</td>
<td>9.60 (1.62, 2.22)</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>878</td>
<td>802</td>
<td>36.26 (1.42, 2.09)</td>
<td>36.26</td>
<td>36.26 (1.42, 2.09)</td>
</tr>
<tr>
<td>Total events: 487 (Treatment), 329 (Control)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for heterogeneity: $\chi^2 = 3.47, df = 4 (P = 0.48), P = 0%$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 5.51 (P &lt; 0.0001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>2374</td>
<td>2243</td>
<td>100.00 (1.24, 1.57)</td>
<td>100.00</td>
<td>100.00 (1.24, 1.57)</td>
</tr>
<tr>
<td>Total events: 1055 (Treatment), 817 (Control)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for heterogeneity: $\chi^2 = 31.85, df = 12 (P = 0.001), I^2 = 62.3%$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 5.45 (P &lt; 0.0001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ITT indicates intent-to-treat; LOCF, last observation carried forward; OR, odds ratio; 95% CI, the 95% confidence interval.

Am J Geriatr Psychiatry 16:7, July 2008 565
Further examination of these and other potential moderators of response require data that often are not published and would benefit from analyses at the individual patient level. Because overall drug effects seem modest, determination of the factors moderating response is critical to decisions about when to use antidepressants and when to consider other treatments.

We used a fixed-effects model because the trials’ had similar designs, inclusion criteria, interventions, and outcomes. One way to approach the heterogeneity of outcomes is with a random effects model to allow that the trials may be estimating different, yet related, treatment effects. This more conservative approach produced results similar to the fixed-effects model (OR = 1.38 versus 1.40) but with wider CI (the 95% CI increased from 1.24–1.57 to 1.12–1.69, and a change in the p value from p <0.001 to p = 0.002).

Although our search excluded trials of medications not marketed in the US, we found no trials of such agents that met our inclusion criteria. Three small placebo-controlled trials in older depressed patients—one with mianserin and two with nomifensine—were published between 1982 and 1984, but they did not report ITT response or remission rates using a structured scale. Four other placebo-controlled trials with drugs approved for depression outside the US (moclobemide, fluvoxamine, and reboxetine) were excluded because of inclusion of inpatients or medically ill populations. No placebo-controlled trials of milnacipram or tianeptine in elderly depressed patients were found. We considered all non-TCA antidepressants as “second-generation” agents, but no trials of MAOIs, trazodone, mirtazapine, or nefazodone were found that met the selection criteria.

This systematic review and meta-analysis contrasts with a recent review of late-life depression studies that included a broad array of depressive disorders, diagnoses, and patients, and various drugs and psychotherapies. That review included studies of alternative drugs and supplements (e.g., acetyl-L-carnitine, alprazolam, tryptophan), patients with depressive symptoms secondary to specific medical conditions (e.g., stroke, obstructive lung disease, Alzheimer disease), and trials with younger, middle-aged patients, (e.g., mean ages of 60). By comparison, we focused on unipolar, major depression trials and 6 of the 10 trials that we included were not included in that review.

This analysis has limitations. The trials included were typical outpatient clinical trials that excluded patients with unstable medical illness and patients in residential settings, and thus our findings should not be generalized to frail elderly. Most of the patients included in the trials were within the range of 60–80 years of age and thus, the findings may not pertain to those of advanced age. All the trials were sponsored by the manufacturer of one of the drugs studied, and although we performed a systematic review, it is possible that there are other trials that have been performed but not reported. Finally, the number of trials included in the analysis is small.

In summary, second-generation antidepressants seem effective for late-life depression but the magnitude of the effect is small. Duration of treatment may be important in that response rates and the specific effects of antidepressant medication (versus placebo) increase with time. The clinical decision to employ antidepressants will need to weigh these modest benefits during acute treatment with more robust effects on prevention of relapse or recurrence and potential safety issues such as increased risks of bleeding, hyponatremia, and decreased bone density and fractures. Finally, the heterogeneity of responses suggests substantial variability among individuals and that determination of moderators or predictors of response is especially important in order to identify patients for whom drug treatment is likely to be effective and those not likely to benefit. The effort to test these hypotheses and to examine moderating variables will benefit from analysis of individual patient data from these trials.

Dr. Nelson conducted the initial search. Drs. Nelson and Schneider performed the trial selection and data checking. All authors participated in analysis of the data and preparation of the manuscript.

The authors received no financial support for this meta-analysis.

This meta-analysis was presented at the Annual Meeting of the AAGP, New Orleans, March 1–4, 2007.
References