



Published in final edited form as:

Arch Gen Psychiatry. 2012 June ; 69(6): 572–579. doi:10.1001/archgenpsychiatry.2011.2044.

Who Benefits from Antidepressants?:

Synthesis of 6-Week Patient-Level Outcomes from Double-Blind Placebo Controlled Randomized Trials of Fluoxetine and Venlafaxine

Robert D. Gibbons^{1,2}, Kwan Hur^{1,3}, C. Hendricks Brown^{1,4}, John M. Davis⁵, and J. John Mann⁶

¹Center for Health Statistics, University of Chicago, 5841 S. Maryland Avenue, MC 2007 office W260, Chicago IL 60637

²Departments of Medicine and Health Studies, University of Chicago, 5841 S. Maryland Avenue, MC 2007 office W260, Chicago IL 60637

³Center for Medication Safety, Pharmacy Benefit Management Services, Hines VA Hospital, Hines, IL 60141

⁴Prevention Science and Methodology Group, Center for Family Studies, Department of Epidemiology and Public Health, University of Miami, 1425 N.W. 10th Avenue, Miami, Florida 33136

⁵Department of Psychiatry, University of Illinois at Chicago, Rooms 455-457 (MC 912), 1601 W. Taylor, Chicago, IL 60612, USA

⁶Department of Molecular Imaging and Neuropathology, New York State Psychiatric Institute, Department of Psychiatry, Columbia University College of Physicians & Surgeons, 1051 Riverside Drive, New York, NY 10032, USA

Abstract

Context—Some meta-analyses suggest that efficacy of antidepressants for major depression is over-stated and limited to severe depression.

Objective—To determine short-term efficacy of antidepressants for treating major depression in youth, adults and geriatric populations.

Design—Reanalysis of all intent-to-treat person-level longitudinal data during the first 6 weeks of treatment of major depressive disorder from 12 adult, 4 geriatric and 4 youth RCTs of fluoxetine and 21 adult trials of venlafaxine.

Setting—All sponsor conducted RCTs of fluoxetine and venlafaxine.

Main Outcome Measures—Children's Depression Rating Scale (youth), the Hamilton Depression Rating Scale (adult and geriatric) and estimated response and remission rates at 6 weeks.

Patients—Fluoxetine – 2635 adult, 960 geriatric and 708 youth. Venlafaxine - 2421 IR and 2461 ER adult.

Results—Patients in all age and drug groups had significantly greater improvement relative to placebo controls. Differential rate of improvement was largest for adult fluoxetine patients (35%

greater than placebo). Youth had the largest treated versus control difference in response rates (24.1%) and remission rates (30.1%), with adult differences generally in the 15%–22% range. Geriatric patients had the smallest drug-placebo differences, 19% greater rate of improvement, 10% for response and 7% for remission. Venlafaxine IR produced larger effects than ER. Baseline severity could not be shown to affect symptom reduction.

Conclusions—This is the first research synthesis in this area to use complete longitudinal person-level data from a large set of published and unpublished studies. The results do not support previous findings that antidepressants show little benefit except for severe depression. The antidepressants fluoxetine and venlafaxine are efficacious for major depression, in all age groups although more so in youth and adults compared with geriatric patients. Baseline severity was not significantly related to degree of treatment advantage over placebo.

Introduction

Recent reports suggest that efficacy of antidepressant medications versus placebo may be overstated, due to publication bias (1) and less efficacy for mildly depressed patients (2,3). For example (1), of 74 FDA-registered randomized controlled trials (RCTs) involving 12 antidepressants in 12,564 patients, 94% of published trials were positive whereas only 51% of all FDA registered studies were positive. A meta-analysis (2) of 35 RCTs of four antidepressants (fluoxetine, venlafaxine, nefazodone, and paroxetine) found that only studies with higher average baseline severity achieved the putative clinically significant 3-point difference in Hamilton Depression Scale (HAM-D) scores (4). A study (3) of individual level data from six RCTs comparing paroxetine or imipramine to placebo concluded there was benefit only for those with very severe depression.

Questions regarding these studies have been raised (5). First, the RCT data may not generalize to patients in the “real world,” who could be switched to another medication if unresponsive. The NIMH-funded STAR*D trial found a 67% remission rate for patients who finished all four phases of the trial. Second, a recent meta-analysis (6) of new-generation antidepressants (56 RCTs in 7,334 patients) failed to detect an association between the trial’s average baseline severity and treatment response. Third, there is a lack of clinically informative outcomes such as the percentage of patients experiencing response or remission.

We question the meaning of a relationship between average study-level initial severity and treatment response (7). Patient level data are required to draw this type of inference as conducted in (3). We also question the use of so-called “vote counting” methods employed in (1) where a simple tally of “positive” studies is used to draw an inference to overall efficacy of a treatment. Our work extends beyond (3) by including a much larger number of trials, accounting for heterogeneity in growth curves within and across studies, and better handling of dropouts (8–10).

We obtained complete longitudinal patient records for RCTs of fluoxetine (a widely-used, selective serotonin reuptake inhibitor - SSRI) conducted by Lilly, the NIMH Treatment for Adolescent Depression Study (TADS) of fluoxetine in youth, and adult studies for venlafaxine (a widely-used serotonin-norepinephrine reuptake inhibitor - SNRI) conducted by Wyeth. These patient level data allow us to: (a) examine associations between treatment response and baseline severity measured at the patient level; (b) use all available longitudinal data from each subject; (c) fit models with less restrictive assumptions regarding missing data; and (d) minimize the effects of selection/publication bias by including nearly all of the placebo controlled depression RCTs of fluoxetine and venlafaxine.

Methods

Study Data

All trials were double-blind, placebo controlled RCTs. For fluoxetine, we re-analyzed studies that included 30 or more patients and used the HAM-D for adults and geriatric patients, or the Childhood Depression Rating Scale-Revised (CDRS-R) for youth. The only trial exclusions were: one adult study that did not use HAM-D, one study judged to be invalid, and one youth study was excluded because it did not use CDRS-R. Fluoxetine trial data from TADS (11) were obtained from NIMH; individual level data for the remaining 12 adult studies, 4 geriatric studies, and 3 youth studies were obtained from Eli Lilly. We obtained patient level data from Wyeth for all available adult venlafaxine RCTs (11 IR and 10 ER).

For fluoxetine, we analyzed data from 12 adult studies with 2,635 patients and 14,048 measurements; 4 geriatric studies with 960 patients and 5,209 measurements; and 4 youth studies with 708 patients and 2,536 measurements. For the adult trials, 11 were outpatient and one had both inpatient and outpatient settings. Inclusion required a diagnosis of depression closely resembling MDD, and the modal trial dosage was 20 mg/d of fluoxetine (range, 20–80 mg). The geriatric trials included 3 outpatient and one inpatient trial, all subjects had depression (similar to MDD), and were over 60 years of age. The modal trial dosage was 20 mg/d (range, 10–30 mg). The youth data consisted of 3 outpatient and one inpatient trial, all subjects had depression (similar to MDD), and ages ranged from 7–18 yr. The modal trial dosage was 20 mg/d (range 10–40 mg).

The TADS study was a 2×2 factorial design of fluoxetine and cognitive behavior therapy (CBT) for the treatment of adolescent major depressive disorder. The study involved 439 patients between the ages of 12 and 17 randomly assigned to the four treatment conditions. Our analysis included only placebo and fluoxetine arms. Study dosages were 10–40 mg/d. Doses of placebo and fluoxetine began at a starting dose of 10 mg/d, which was then increased to 20 mg/d at week 1 and if necessary, to a maximum of 40 mg/d by week 8. Modal dosages were not available for the venlafaxine trials; however, ranges are reported in Table 1.

For venlafaxine IR, there were 11 adult studies with 2,421 patients and 10,634 measurements and for ER; there were 10 adult studies with 2,461 patients and 12,481 measurements. The majority of the studies were outpatient (IR: 2 inpatient and 9 outpatient, ER: 10 outpatient). Dosages were in the range of 25–375 mg (modal range 75–150 mg). Table 1 provides a summary of the trials.

Statistical Methods

Analyses were conducted using SuperMix (12). Data were analyzed using a three-level mixed-effects linear regression model (10). Level 1 represents the measurement occasion, level 2 the patient, and level 3 the study. An overall analysis was performed for both drugs (fluoxetine and venlafaxine IR and ER) in adults and geriatrics. Youth were not included in the combined analysis because they were assessed with the CDRS. Separate analyses were performed for fluoxetine adult, geriatric, and youth trials, and venlafaxine IR and ER adult trials. The intercept and slope of the time trends were random effects at both patient and study levels, allowing time-trends to vary across patients and studies. Heterogeneity of treatment effect was tested by including a random treatment by time interaction. Time was number of days from treatment initiation. The primary effect of interest was change in slope between treatment and control over 6 weeks. Six weeks was selected because it was the minimum trial duration and therefore we were able to perform an analysis in which study and length of treatment were unconfounded. We report the marginal maximum likelihood

estimate (MMLE), and standard error (SE) of the difference in HAM-D or CDRS scores at 6 weeks between treated and control groups. As a sensitivity analysis we analyzed scheduled weekly visit data (i.e., intended week of the visit instead of the actual day of the visit). We evaluated models with non-linear time trends; however, the linear model provided the best overall fit to the data through six weeks. To test for dependence of treatment effect on initial severity we dichotomized baseline severity (HAM-D>20 (13); CDRS-R>60 (14)) at the subject level and included interactions with treatment and time effects, with the estimate of interest being the 3-way severity by treatment by time interaction. The data were also analyzed using a continuous baseline severity score. These analyses included all two and three-way interactions to determine if baseline severity moderated the relationship between treatment and response. All analyses were adjusted for main effects of age and sex, which decreases variability, but does not determine if age and sex are treatment moderators. Age-specific analyses (youth, adult, geriatric) address the question of differential treatment effects across the lifespan.

Empirical Bayes estimates were used to estimate response and remission at the end of six weeks for each patient, and compared using a mixed-effects logistic regression model adjusting for study. As a sensitivity analysis, the observed baseline and week six HAM-D (or CDRS-R) scores were used if available. 52.2% of all subjects had a score during week 6 and 21.4% had a score on day 42. The observed day 42 scores were used in the sensitivity analysis. For adults and geriatric patients, response was a 50% reduction in HAM-D score at week six, and remission was HAM-D<8. For youth, response was a 50% reduction in CDRS-R score at week six and remission was CDRS-R<28. The number needed to treat (NNT) to obtain a single additional remission or response in the treatment arm relative to the placebo arm was also reported.

Results

Overall Analysis for Fluoxetine and Venlafaxine IR and ER in Adults and Geriatrics

The estimated average rate of change over 6 weeks was -11.82 HAM-D units for drug versus -9.26 HAM-D units for placebo (MMLE = -2.55 , SE = 0.20 , $p<0.0001$), indicating 28% greater improvement for drug. Analyses based on weekly data yielded virtually identical results. Estimated linear time trends and observed daily mean scores are presented in Figure 1. Variation in the treatment effect over studies was not statistically significant (SD = 0.16 , $p=0.06$).

The estimated response rate for drug was 58.4% versus 39.9% for placebo (OR=2.11, 95% CI 1.93–2.31, $p<0.001$, NNT=5.41). Similar results were obtained using available observed HAM-D scores (59.1% versus 41.9%, OR=2.00, 95% CI 1.83–2.19, $p<0.001$, NNT=5.82). Remission rates were 43.0% versus 29.3% for drug and placebo respectively (OR=1.82, 95% CI 1.66–2.00, $p<0.001$, NNT=7.30) and the sensitivity analysis revealed similar results (42.4% versus 28.9%; OR=1.81, 95% CI 1.65–1.99, $p<0.001$, NNT=7.39).

No effect of baseline severity on treatment efficacy was found for either the dichotomous ($p=0.27$) or the continuous ($p=0.10$) baseline severity measures. For patients with low severity, rate of change in symptoms over 6 weeks was -9.40 for drug versus -7.20 for placebo. For high severity patients the rate of change was -12.85 for drug versus -10.07 for placebo. The estimated difference was 2.20, 95% CI 1.65–2.76 for low severity and 2.78, 95% CI 2.26–3.29 for high severity. Estimated response rates were 54.8% vs. 37.3% (difference of 17.5%) for treated vs. placebo low severity patients and 57.7% vs. 40.5% (difference of 17.2%) for high severity patients. Estimated remission rates were 49.9% vs. 36.6% (difference of 13.3%) for treated vs. placebo low severity patients and 37.8% vs. 25.1% (difference of 12.7%) for high severity patients.

Adult Studies of Fluoxetine

The estimated average rate of change over 6 weeks was -10.12 HAM-D units for fluoxetine and 7.52 HAM-D units for placebo (MMLE = -2.60 , SE = 0.34 , $p < 0.0001$), indicating 35% greater improvement for fluoxetine.

The estimated response rate for fluoxetine was 55.1% versus 33.7% for placebo (OR=2.41, 95% CI 1.93–3.01, $p < 0.001$, NNT=4.69). Remission rates were 45.8% versus 30.2% for fluoxetine and placebo respectively (OR=1.96, 95% CI 1.66–2.31, $p < 0.001$, NNT=6.40).

No effect of baseline severity on treatment efficacy was found ($p = 0.14$). For patients with low severity, the rate of change in symptoms over 6 weeks was -7.98 versus -6.26 for placebo. For high severity patients the rate of change for fluoxetine was -11.72 versus -8.32 for placebo. The estimated difference was 1.68, 95% CI 0.77–2.59 for low severity and 3.40, 95% CI 2.49–4.31 for high severity. Estimated response rates were 50.0% vs. 36.4% (difference of 13.6%) for treated vs. placebo low severity patients and 56.2% vs. 32.7% (difference of 23.5%) for high severity patients. Estimated remission rates were 51.7% vs. 41.2% (difference of 10.5%) for treated vs. placebo low severity patients and 37.8% vs. 22.1% (difference of 15.7%) for high severity patients. Neither response rates nor rates of improvement differed statistically between low and high severity groups.

Adult Studies of Venlafaxine ER

The estimated average rate of change over 6 weeks was -12.39 HAM-D units for venlafaxine ER and -10.21 HAM-D units for placebo (MMLE = -2.18 , SE = 0.38 , $p < 0.0001$), indicating 22% greater improvement for venlafaxine ER.

The estimated response rate for venlafaxine ER was 60.5% versus 45.4% for placebo (OR=1.92, 95% CI 1.61–2.29, $p < 0.001$, NNT=6.63). Remission rates were 41.5% versus 29.5% for venlafaxine ER and placebo respectively (OR=1.75, 95% CI 1.46–2.09, $p < 0.001$, NNT=8.31).

No effect of baseline severity on treatment efficacy was found ($p = 0.94$). For low severity patients, the rate of change was -9.58 for venlafaxine ER versus -7.10 for placebo. For high severity the rate of change was -12.81 for venlafaxine ER versus -10.58 for placebo.

The estimated difference was 2.48, 95% CI 0.91–4.05 for low severity and 2.23, 95% CI 1.40–3.05 for high severity. Estimated response rates were 53.1% vs. 36.6% (difference of 16.5%) for treated vs. placebo low severity patients and 59.9% vs. 45.5% (difference of 14.4%) for high severity patients. Estimated remission rates were 52.1% vs. 35.2% (difference of 16.9%) for treated vs. placebo low severity patients and 40.0% vs. 28.6% (difference of 11.4%) for high severity patients.

Adult Studies of Venlafaxine IR

For venlafaxine IR, the rate of change over 6 weeks was -14.32 HAM-D units for venlafaxine IR and -10.71 HAM-D units for placebo (MMLE = -3.61 , SE = 0.42 , $p < 0.0001$), indicating 34% greater improvement for venlafaxine IR.

The estimated response rate for venlafaxine IR was 67.2% versus 45.2% for placebo (OR=3.16, 95% CI 2.52–3.97, $p < 0.001$, NNT=4.53). Remission rates were 47.1% versus 32.4% for venlafaxine IR and placebo respectively (OR=2.24, 95% CI 1.82–2.75, $p < 0.001$, NNT=6.83).

No effect of baseline severity on treatment efficacy was found ($p = 0.29$). For low severity patients, the rate of change was -10.08 for venlafaxine IR versus -7.69 for placebo. For

high severity patients the rate of change was -14.87 for venlafaxine IR versus -10.88 for placebo. The estimated difference was 2.39 , 95% CI $0.33-4.46$ for low severity and 3.99 , 95% CI $3.08-4.90$ for high severity. Estimated response rates were 64.7% vs. 43.9% (difference of 20.8%) for treated vs. placebo low severity patients and 66.5% vs. 43.0% (difference of 23.5%) for high severity patients. Estimated remission rates were 57.0% vs. 43.3% (difference of 13.7%) for treated vs. placebo low severity patients and 45.3% vs. 29.4% (difference of 15.9%) for high severity patients.

Effect of Age Group on Treatment Response

Geriatric Studies of Fluoxetine—For fluoxetine, the estimated average rate of change over 6 weeks was -7.48 HAM-D units for placebo and -8.86 HAM-D units for fluoxetine (MMLE = -1.39 , SE = 0.50 , $p < 0.009$), indicating 19% greater improvement for fluoxetine. However, response and remission rates were not significantly different between treated and control conditions.

The estimated response rate for fluoxetine was 37.3% versus 27.4% for placebo (OR= 1.42 , 95% CI $0.92-2.18$, $p = 0.115$, NNT= 16.95). Remission rates were 26.5% versus 20.0% for fluoxetine and placebo respectively (OR= 1.26 , 95% CI $0.78-2.03$, $p = 0.337$, NNT= 38.71).

No effect of baseline severity on treatment efficacy was found ($p = 0.95$). For low severity patients, the rate of change was -6.89 for fluoxetine versus -5.42 for placebo. For high severity patients the rate of change for fluoxetine was -9.37 versus -8.02 for placebo. The estimated difference was 1.47 , 95% CI $-0.26-3.20$ for low severity and 1.34 , 95% CI $0.02-2.67$ for high severity. Estimated response rates were 38.1% vs. 26.9% (difference of 11.2%) for treated vs. placebo low severity patients and 37.1% vs. 26.5% (difference of 10.6%) for high severity patients. Estimated remission rates were 38.1% vs. 27.6% (difference of 10.5%) for treated vs. placebo low severity patients and 22.3% vs. 16.4% (difference of 5.9%) for high severity patients.

Youth Studies—The estimated average rate of change over 6 weeks was -15.96 CDRS-R units for placebo and -20.58 CDRS-R units for fluoxetine (MMLE = -4.62 , SE = 1.26 , $p < 0.0001$), indicating 30% greater improvement for fluoxetine.

The estimated response rate for fluoxetine was 29.8% versus 5.7% for placebo (OR= 6.66 , 95% CI $3.07-14.48$, $p < 0.001$, NNT= 4.16). Remission rates were 46.6% versus 16.5% for fluoxetine and placebo respectively (OR= 4.23 , 95% CI $2.64-6.77$, $p < 0.001$, NNT= 3.33). The finding of higher remission than response rates questions the validity of the CDRS-R remission threshold score of 28.

No effect of baseline severity on treatment efficacy was found ($p = 0.90$). For low severity patients, the rate of change in symptoms was -17.60 for fluoxetine versus -12.56 for placebo. For high severity patients the rate of change was -28.86 for fluoxetine versus -24.40 for placebo. The estimated difference was 5.04 , 95% CI $2.56-7.52$ for low severity and 4.45 , 95% CI $-0.58-9.49$ for high severity. Estimated response rates were 23.0% vs. 3.2% (difference of 19.8%) for treated vs. placebo low severity patients and 40.2% vs. 17.2% (difference of 23.0%) for high severity patients. Estimated remission rates were 54.1% vs. 19.4% (difference of 20.6%) for treated vs. placebo low severity patients and 28.9% vs. 7.5% (difference of 21.4%) for high severity patients.

Discussion

We examined symptom trajectories through 6 weeks from all double blind placebo controlled RCTs with fluoxetine and venlafaxine that were conducted by the sponsors.

Statistically and clinically significant benefits of treatment were found. Based on relative change in slopes, remission, response rates, and NNT, treatment effect was largest for youth followed by adults, and more limited for geriatrics. Similar results were found for fluoxetine and venlafaxine.

While average differences at 6 weeks are relatively small, they translate into clinically significant differences in response and remission rates. In adults treated with fluoxetine, 55.1% of treated patients achieved response (50% reduction in severity) compared with only 33.7% of controls and is similar to previous causal inference (growth mixture modeling) findings for fluoxetine and imipramine (15). From a public health perspective this is an enormous difference and indicates that for every 5 treated patients an additional patient treated with fluoxetine will respond. Similarly, remission rates were 45.8% for treated patients but only 30.2% for controls. Even stronger results were observed for children. In youth studies, 29.8% of treated children responded whereas only 5.7% of children on placebo responded. Similarly, remission rates were 46.6% for treated children but only 16.5% for controls. These rates translate into an additional child treated with fluoxetine responding and remitting for every four children treated. The higher rates of remission suggest that the remission criterion (CDRS-R=28) should be re-evaluated.

By contrast, we found statistically significant (for HAM-D scores but not remission and response rates) but much less clinically significant effects for geriatrics. Response rates were 37.3% versus 27.4% translating to one additional patient responding on fluoxetine for every 17 patients treated. Remission rates were 26.5% versus 20.0%, which translates into one additional patient remitting on fluoxetine for every 39 patients treated. The efficacy of antidepressant treatment in geriatric patients should be studied in greater detail based on these findings. There may be a biological explanation for the age effect on response rates since both neuroendocrine challenge studies and receptor imaging studies report poorer antidepressant responses in depressed patients with more pronounced serotonin abnormality (16–17). Serotonin function declines with age, potentially increasing the proportion of such patients in geriatric studies.

Venlafaxine produced similar results to fluoxetine, suggesting these results are not specific to fluoxetine. Increased efficacy for the IR versus ER formulations was observed and should be further studied.

Perhaps most importantly these findings illustrate that relatively small overall mean differences can translate into relatively large patient-level differences in clinically interpretable and meaningful endpoints such as response and remission. Statistically, these small changes in the mean of the distribution can often translate into much larger effects in the tails of the distribution.

Most studies were designed for achieving regulatory approval and do not demonstrate the maximum effect that a drug can produce. Some studies were as short as 6 weeks in duration, whereas the maximum effect during an acute treatment episode is likely 12 weeks or longer. Few well-controlled studies, other than the long-term maintenance study of Frank and colleagues (18) have documented response rates for extended treatment with a single effective antidepressant. In that study remission rate was 82%, with 75% achieving remission by 140 days (17). For fluoxetine, 23% of patients who were unimproved at 8 weeks showed full remission at 12 weeks (19).

The findings of this study shed light on meta-analytic results that related average study-level initial severity to the magnitude of treatment response. When examined at the patient level, baseline severity did not moderate treatment response for any endpoint, age group, or drug. Overall response rates were lower for geriatrics than adults but did not vary by baseline

severity. For children, response rates were lower overall compared to adults; however, here there were substantial differences between low and high baseline severity groups for both treated and control patients (i.e., not a treatment related effect).

Results of this study raise serious questions regarding the results of meta-analyses that are now so prevalent in guiding medical decisions. In addition to the obvious issues related to publication bias (1), the use of average endpoints gleaned from studies that use a variety of different approaches to handling missing data (e.g., LOCF or completer analyses), and the loss of intermediate longitudinal measurements and their associated contribution to the overall estimate of variability, can yield biased results. Reliance on meta-regression to examine relationships that exist at the person-level but are analyzed at the study level, can lead to erroneous conclusions that are not supported when all available person-level data are available. We note; however, that the approach to research synthesis taken in this paper requires that all studies use a common endpoint. When different studies have used different endpoints, then this is exactly the type of problem that meta-analysis was designed for, and for which it should be used. In these cases; however, one must take great care to use a well chosen effect size (not a mean difference, for example), that is both statistically and clinically meaningful.

There are several limitations of the present study. First, we considered only two antidepressant medications, and other antidepressants may produce different effects. Indeed, fluoxetine is the only antidepressant that is approved for the treatment of childhood depression and it was the only antidepressant that we studied in children. Second, there were only four youth trials and we must therefore interpret the estimated efficacy observed in these trials with caution. However, the rather impressive effects on clinically interpretable outcomes such as response and remission indicate the clinical benefit that children may receive with careful pharmacologic treatment. These findings should also favor reconsideration of the risk benefit equation that led to the black box warning for suicidal thinking and antidepressants in children (20). Third, a similar note of caution is in order for the results for the four geriatric studies, where some statistically significant but more marginal clinically significant results were observed. Fourth, the reported findings are limited to the first 6 weeks of study. Results may differ for long-term outcomes, and may be stronger as placebo benefit may degrade over time (19).

Fifth, it is possible that some selection bias remains even though our synthesis included all studies conducted by the sponsors and was not limited to the subset of studies that were in the published literature. Sixth, our study used industry sponsored studies, which were designed to demonstrate efficacy and may have enrolled patients that may not have been representative of the patients seeking treatment for depression. However, two of the four youth studies were academic studies (TADs and the study X065). Since the largest effect of treatment was seen for youth, and these studies are a mix of industry and academic studies, it seems unlikely that reliance on industry-sponsored studies produced biased results. Seventh, the majority of the studies had placebo lead-in periods that are designed to eliminate early placebo responders. However, an analysis of 75 RCTs involving antidepressants and placebo from 1981–2000 (21) found that the use of a placebo lead-in period did not relate to the response rate in the placebo group ($p=.73$). Like us they also found that baseline severity did not predict placebo response. We note however, that while this analysis did not find any effect of a prospective lead-in on placebo response rates, and it did not find a relationship to baseline severity, it is possible that the method of analysis (analyzing response rates as opposed to absolute magnitude of change) may have missed meaningful effects of the lead-in.

To determine if we included the majority of placebo controlled depression studies of fluoxetine, we reviewed the published literature on placebo-controlled RCTs of fluoxetine in the acute treatment of major depressive disorder that met the following criteria: 1) not sponsored by a pharmaceutical company; 2) not associated with a specific medical illness (e.g., post myocardial infarction, AIDS); 3) not associated with comorbid substance abuse (including alcohol); 4) not associated with a specific diagnosable comorbid psychiatric disorder; 5) used the Hamilton Depression Scale; 6) had a minimum enrollment of 30 patients. Knowledge Finder™ (Aries System; North Andover, MA) was used to search the PubMed database from 1966 through October 31, 2010. The Boolean search option, with word variants, was used to search “placebo controlled trials of fluoxetine in major depression”. The search returned 329 references. Titles and abstracts of these references were reviewed to find articles potentially meeting the above criteria (n=13) as well as articles that were reviews or meta-analyses (n=7). The reference lists for the reviews and meta-analyses were inspected for additional articles potentially meeting the above criteria (n=1). Following these two manual reviews, reprints of candidate articles were obtained and reviewed for the 14 candidate articles. Two articles fulfilled the above criteria (22, 23). The first study (22) was restricted exclusively to patients meeting the Columbia criteria for atypical depression (while meeting criteria for major depressive disorder). This study was partially funded by Lilly and their data were not available to us. The second study (23) was a small study conducted in Brazil comparing St. John’s wort (n=20), with fluoxetine (n=20) and placebo (n=26) and was partially funded by the company supplying the St. John’s wort. This literature search confirms that few if any academic studies of fluoxetine in the treatment of adult depression were conducted and that our data represent the majority of available published (12 studies) and unpublished (8 studies) RCTs.

In conclusion, a detailed research synthesis using patient-level longitudinal data from all available youth, adult, and geriatric placebo controlled RCTs of fluoxetine conducted by the sponsor reveals consistent statistically significant benefits of treatment, the magnitude of which was greatest in youth and smallest in geriatric subpopulations (where differences in remission and response rates did not reach statistical significance). Analyses of venlafaxine RCTs confirmed the results for the efficacy of antidepressant treatment in adults. Baseline severity did not moderate the effect of treatment. Similar re-analyses should be conducted with other newer antidepressant medications to confirm these findings. This study also highlights many of the limitations of meta-analysis that combine evidence from multiple RCTs (e.g., meta-regression of study-level characteristics that exhibit inter-individual variability, inconsistent and potentially biased handling of missing data, etc.) and further highlights advantages of a more complete person-level analysis when such data are available, and increases the need for caution regarding interpretation of meta-analytic results when person-level data are not available.

Acknowledgments

This work was supported by NIMH grants MH062185 (JJM) and R56 MH078580 and MH8012201 (RDG and CHB), MH040859 (CHB), and AHRQ grant 1U18HS016973 (RDG). Dr. Gibbons has served as an expert witness for the U.S. Department of Justice, Wyeth and Pfizer Pharmaceuticals in cases related to antidepressants and anticonvulsants and suicide. Dr. Mann has received research support from Glaxo Smith Kline and Novartis. Dr. Brown directed a suicide prevention program at the University of South Florida that received funding from JDS Pharmaceuticals. Data were supplied by NIMH (TADS study), Wyeth, and Eli Lilly. Dr. Gibbons had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med.* 2008; 358:252–260. [PubMed: 18199864]

2. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med.* 2008; 5:e45. [PubMed: 18303940]
3. Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, Fawcett J. Antidepressant drug effects and depression severity. A patient-level meta-analysis. *JAMA.* 2010; 303:47–53. [PubMed: 20051569]
4. National Institute for Clinical Excellence. Depression: management of depression in primary and secondary care. Clinical practice guideline No 23. London: National Institute for Clinical Excellence; 2004. p. 670
5. Mathew SJ, Charney DS. Publication bias and the efficacy of antidepressants. *American Journal of Psychiatry.* 2009; 166:140–145. [PubMed: 19188290]
6. Melander H, Salmonson T, Abadie E, van Zwieten-Boot B. A regulatory apologia: a review of placebo-controlled studies in regulatory submissions of new-generation antidepressants. *Eur Neuropsychopharmacol.* 2008; 18:623–627. [PubMed: 18621509]
7. Brown CH, Wang W, Sandler I. Examining how context changes intervention impact: The use of effect sizes in multilevel mixture meta-analysis. *Child Develop Perspectives.* 2008; 2:198–205.
8. Lavori PW, Brown CH, Duan N, Gibbons RD, Greenhouse J. Missing data in longitudinal clinical trials - Part A: Design and conceptual issues. *Psychiatric Annals.* 2008; 38:784–792. [PubMed: 19756211]
9. Siddiqui O, Hung HM, O'Neill R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. *J Biopharm Stat.* 2009; 19:227–246. [PubMed: 19212876]
10. Hedeker, D.; Gibbons, RD. *Longitudinal Data Analysis.* New York: John Wiley & Sons; 2006.
11. Treatment for adolescents with depression study team. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for adolescents with depression study (TADS) randomized controlled trial. *JAMA.* 2004; 292:807–820. [PubMed: 15315995]
12. Hedeker, D.; Gibbons, RD.; Du Toit, SHC.; Patterson, D. *SuperMix - A program for mixed-effects regression models.* Scientific Software International; Chicago: 2008.
13. Elkin I, Gibbons RD, Shea MT, Sotsky SM, Watkins JT, Pilkonis PA, Hedeker D. Initial severity and differential outcome in the NIMH treatment of depression collaborative research program. *Journal of Clinical and Consulting Psychology.* 1995; 63:841–847.
14. March, JS. *Independent Evaluator Manual: Treatment for Adolescents with Depression Study (TADS), Final Version 4.1.* Duke University Medical Center; 2005.
15. Muthén B, Brown CH. Estimating drug effects in the presence of placebo response: Causal inference using growth mixture modeling. *Statistics in Medicine.* 2009; 28:3363–3385. [PubMed: 19731223]
16. Malone KM, Thase ME, Mieczkowski TA, Myers JE, Perel JM, Cooper TB, Mann JJ. Fenfluramine challenge test as a predictor of outcome in major depression. *Psychopharmacology Bulletin.* 1993; 29(2):155–161. [PubMed: 7507256]
17. Parsey RV, Olvet DM, Oquendo MA, Huang Y, Ogden RT, Mann JJ. Higher 5-HT_{1A} receptor binding potential during a major depressive episode predicts poor treatment response: Preliminary data from a naturalistic study. *Neuropsychopharmacology.* 2006; 31:1745–1749. [PubMed: 16395308]
18. Frank E, Kupfer DJ, Perel JM, Cornes C, Jarrett DB, Mallinger AG, Thase ME, McEachran AB, Grochocinski VJ. Three-year outcomes for maintenance therapies in recurrent depression. *Arch Gen Psychiatry.* 1990; 47:1093–1099. [PubMed: 2244793]
19. Quitkin FM, Petkova E, McGrath PJ, Taylor B, Beasley C, Stewart J, Amsterdam J, Fava M, Rosenbaum J, Reimherr F, Fawcett J, Chen Y, Klein D. When should a trial of fluoxetine for major depression be declared failed? *American Journal of Psychiatry.* 2003; 160:734–740. [PubMed: 12668363]
20. Bridge JA, Iyengar S, Salary CB, Barbe RP, Birmaher B, Pincus HA, Ren L, Brent DA. Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment. A meta-analysis of randomized controlled trials. *JAMA.* 2007; 297:1683–1696. [PubMed: 17440145]

21. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression. *JAMA*. 2002; 287:1840–1847. [PubMed: 11939870]
22. McGrath PJ, Stewart JW, Janal MN, Petkova E, Quitkin FM, Klein D. A placebo-controlled study of fluoxetine versus imipramine in the acute treatment of atypical depression. *Am J Psychiatry*. 2000; 157:344–350. [PubMed: 10698808]
23. Moreno RA, Teng CT, de Almeida KM, Junior HT. *Hypericum perforatum* versus fluoxetine in the treatment of mild to moderate depression: a randomized double-blind trial in a Brazilian sample. *Rev Bras Psiquiatr*. 2005; 28:29–32. [PubMed: 16612487]

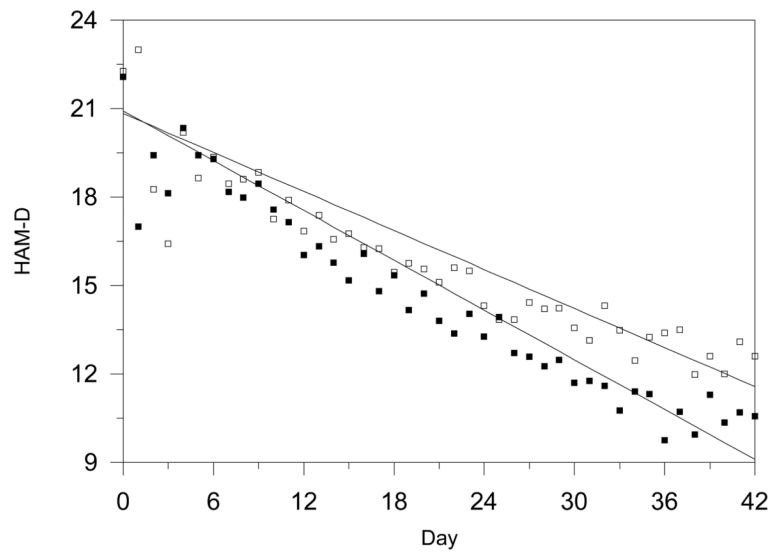


Figure 1. Observed Versus Estimated Depression Severity Time Trends For Drug versus Placebo in 37 Adult and Geriatric Studies

Table 1

Summary of Studies

Study	Study ID	# of Subjects	Treatment		Placebo		Female		Age mean(sd)	Dosing (mg/day)
			n	%	n	%	n	%		
Lilly - Adult	C031	69	37	53.6	32	46.4	46	66.7	38.7 (9.1)	20
	HCAC	110	55	50.0	55	50.0	77	70.0	40.2 (12.5)	20-80
	HCAF	475	244	51.4	231	48.6	311	65.5	40.3 (12.3)	40-80
	HCCH	90	45	50.0	45	50.0	45	50.0	41.8 (12.0)	40-80
	HCCP	733	626	85.4	107	14.6	421	57.4	39.4 (11.4)	20, 40, 60
	HCDD	363	285	78.5	78	21.5	221	60.9	38.9 (11.1)	5, 20, 40
	HCFP	89	46	51.7	43	48.3	63	70.8	40.4 (10.4)	20
	HLAB	166	93	56.0	73	44.0	86	51.8	43.2 (9.6)	20
	HMAQA	103	33	32.0	70	68.0	67	65.0	40.8 (12.3)	20
	HMAQB	112	37	33.0	75	67.0	73	65.2	40.8 (12.1)	20
Lilly - Geriatric	HZAA	128	63	49.2	65	50.8	76	59.4	40.0 (10.1)	10-20
	LUAB	197	63	32.0	134	68.0	128	65.0	40.0 (11.2)	20
	HCCQ	90	59	65.6	31	34.4	64	71.1	73.9 (6.1)	20, 40
	HCFE	670	335	50.0	335	50.0	366	54.6	67.7 (5.7)	20
	HCIR	121	57	47.1	64	52.9	88	72.7	83.7 (5.1)	10-30
	HC55	79	39	49.4	40	50.6	59	74.7	80.7 (6.2)	20
	HCJE	219	109	49.8	110	50.2	108	49.3	12.7 (2.6)	10-20
	LYAQ	172	127	73.8	45	26.2	48	27.9	11.3 (2.6)	10-40
	X065	96	48	50.0	48	50.0	44	45.8	12.8 (2.7)	20
	TADS	014	221	109	49.3	112	50.7	118	53.4	14.6 (1.6)
Wyeth - ER	105	198	102	51.5	96	48.5	100	50.5	71.0 (5.1)	37.5-225
	016	189	94	49.7	95	50.3	107	56.6	40.9 (12.9)	37.5-375
	208	197	97	49.2	100	50.8	119	60.4	39.8 (10.5)	75-150
	209	197	95	48.2	102	51.8	121	61.4	40.7 (11.4)	75-225
	211	198	100	50.5	98	49.5	136	68.7	39.7 (11.7)	75-225
	360	246	127	51.6	119	48.4	147	59.8	41.4 (11.5)	75-225

Study	Study ID	# of Subjects	Treatment		Placebo		Female		Age mean(sd)	Dosing (mg/day)
			n	%	n	%	n	%		
	367	248	165	66.5	83	33.5	168	67.7	44.6 (11.7)	75, 150
	402	395	295	74.7	100	25.3	239	60.5	38.8 (13.1)	37.5-300
	414	391	292	74.7	99	25.3	260	66.5	42.2 (13.5)	37.5-300
Wyeth - IR	203	358	266	74.3	92	25.7	139	38.8	43.0 (10.7)	75, 150-225, 300-375
	206	93	46	49.5	47	50.5	79	84.9	65.2 (13.3)	150-375
	208	196	96	49.0	100	51.0	124	63.3	41.1 (11.0)	75-150
	300	108	58	53.7	50	46.3	33	30.6	43.6 (12.0)	150-375
	301	151	73	48.3	78	51.7	104	68.9	41.7 (12.7)	75-225
	302	148	72	48.6	76	51.4	96	64.9	41.7 (11.6)	75-200
	303	159	79	49.7	80	50.3	109	68.6	38.9 (9.8)	75-225
	313	312	234	75.0	78	25.0	210	67.3	38.5 (10.3)	25, 50-75, 150-200
	342	384	286	74.5	98	25.5	245	63.8	41.8 (11.7)	75, 150, 200
	343	205	136	66.3	69	33.7	153	74.6	38.7 (10.5)	150, 225
	372	307	156	50.8	151	49.2	190	61.9	39.9 (10.0)	missing