

## ONLINE FIRST

## Benefits From Antidepressants

## Synthesis of 6-Week Patient-Level Outcomes From Double-blind Placebo-Controlled Randomized Trials of Fluoxetine and Venlafaxine

Robert D. Gibbons, PhD; Kwan Hur, PhD; C. Hendricks Brown, PhD; John M. Davis, MD; J. John Mann, MD

**Context:** Some meta-analyses suggest that efficacy of antidepressants for major depression is overstated and limited to severe depression.

**Objective:** To determine the short-term efficacy of antidepressants for treating major depressive disorder in youth, adult, and geriatric populations.

**Data Sources:** 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 randomized controlled trials of fluoxetine hydrochloride and 21 adult trials of venlafaxine hydrochloride.

**Study Selection:** All sponsor-conducted randomized controlled trials of fluoxetine and venlafaxine.

**Data Extraction:** Children's Depression Rating Scale-Revised scores (youth population), Hamilton Depression Rating Scale scores (adult and geriatric populations), and estimated response and remission rates at 6 weeks were analyzed for 2635 adults, 960 geriatric patients, and 708 youths receiving fluoxetine and for 2421 adults receiving immediate-release venlafaxine and 2461 adults receiving extended-release venlafaxine.

**Data Synthesis:** Patients in all age and drug groups had significantly greater improvement relative to control pa-

tients receiving placebo. The differential rate of improvement was largest for adults receiving fluoxetine (34.6% greater than those receiving placebo). Youths had the largest treated vs control difference in response rates (24.1%) and remission rates (30.1%), with adult differences generally in the 15.6% (remission) to 21.4% (response) range. Geriatric patients had the smallest drug-placebo differences, an 18.5% greater rate of improvement, 9.9% for response and 6.5% for remission. Immediate-release venlafaxine produced larger effects than extended-release venlafaxine. Baseline severity could not be shown to affect symptom reduction.

**Conclusions:** To our knowledge, 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first research synthesis in this area to use complete longitudinal person-level data from a large set of published and unpublished studies.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 depressive disorder in all age groups, although more so in youths and adults compared with geriatric patients. Baseline severity was not significantly related to degree of treatment advantage over placebo.

*Arch Gen Psychiatry.* 2012;69(6):572-579.

Published online March 5, 2012.

doi:10.1001/archgenpsychiatry.2011.2044

Author Affiliations are listed at the end of this article.

**R**ECENT REPORTS SUGGEST that the efficacy of antidepressant medications compared with placebo may be overstated owing to publication bias<sup>1</sup> and that antidepressants have less efficacy for mildly depressed patients.<sup>2,3</sup> For example,<sup>1</sup> of 74 US Food and Drug Administration-registered randomized controlled trials (RCTs) involving 12 antidepressants in 12 564 patients, 94% of published trials were positive, whereas only 51% of all US Food and Drug Administration-registered studies were posi-

tive. A meta-analysis<sup>2</sup> of 35 RCTs of 4 antidepressants (fluoxetine hydrochloride, venlafaxine hydrochloride, nefazodone hydrochloride, and paroxetine hydrochloride) found that only studies with higher

## See also page 580

average baseline severity achieved the putative clinically significant 3-point difference in Hamilton Depression Scale (HAM-D) scores.<sup>4</sup> A study<sup>3</sup> of individual-level data from 6 RCTs comparing parox-

etine or imipramine hydrochloride with placebo concluded that there was benefit only for those with very severe depression.

Questions regarding these studies have been raised.<sup>5</sup> First, the RCT data may not generalize to patients in the real world, who could be switched to another medication if unresponsive. The National Institute of Mental Health–funded STAR\*D trial found a 67% remission rate for patients who finished all 4 phases of the trial. Second, a recent meta-analysis<sup>6</sup> of new-generation antidepressants (7334 patients in 56 RCTs) 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.<sup>7</sup> Patient-level data are required to draw this type of inference as conducted by Fournier et al.<sup>3</sup> We also question the use of so-called vote-counting methods used by Turner et al<sup>1</sup> where a simple tally of positive studies is used to draw an inference to overall efficacy of a treatment. Our work extends beyond that of Fournier et al<sup>3</sup> by including a much larger number of trials, accounting for heterogeneity in growth curves within and across studies, and better handling dropouts.<sup>8-10</sup>

We obtained complete longitudinal patient records for RCTs of fluoxetine (a widely used selective serotonin reuptake inhibitor) conducted by Eli Lilly and Co, the Treatment for Adolescents With Depression Study of fluoxetine in youths by the National Institute of Mental Health, and adult studies for venlafaxine (a widely used serotonin-norepinephrine reuptake inhibitor) conducted by Wyeth. These patient-level data allow us to do the following: (1) examine associations between treatment response and baseline severity measured at the patient level; (2) use all available longitudinal data from each subject; (3) fit models with less restrictive assumptions regarding missing data; and (4) minimize the effects of selection or 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 reanalyzed studies that included 30 or more patients and used the HAM-D for adults and geriatric patients or the Children's Depression Rating Scale–Revised (CDRS-R) for youths. The only trial exclusions were 1 adult study that did not use the HAM-D, 1 study judged to be invalid, and 1 youth study that did not use the CDRS-R. Fluoxetine trial data from the Treatment for Adolescents With Depression Study<sup>11</sup> were obtained from the National Institute of Mental Health; individual-level data for the remaining 12 adult studies, 4 geriatric studies, and 3 youth studies were obtained from Eli Lilly and Co. We obtained patient-level data from Wyeth for all available adult venlafaxine RCTs (11 with venlafaxine immediate release [IR] and 10 with venlafaxine extended release [ER]).

For fluoxetine, we analyzed data from 12 adult studies with 2635 patients and 14 048 measurements; 4 geriatric studies with 960 patients and 5209 measurements; and 4 youth stud-

ies with 708 patients and 2536 measurements. For the adult trials, 11 were outpatient and 1 had both inpatient and outpatient settings. Inclusion required a diagnosis of depression closely resembling major depressive disorder (MDD), and the modal trial dosage was 20 mg/d of fluoxetine hydrochloride (range, 20-80 mg/d). The geriatric trials included 3 outpatient trials and 1 inpatient trial; all subjects had depression (similar to MDD) and were older than 60 years. The modal trial dosage was 20 mg/d (range, 10-30 mg/d). The youth data consisted of 3 outpatient trials and 1 inpatient trial; all subjects had depression (similar to MDD) and were aged 7 to 18 years. The modal trial dosage was 20 mg/d (range, 10-40 mg/d).

The Treatment for Adolescents With Depression Study was a 2 × 2 factorial design of fluoxetine and cognitive behavior therapy for the treatment of MDD in adolescents. The study involved 439 patients aged 12 to 17 years randomly assigned to the 4 treatment conditions. Our analysis included only placebo and fluoxetine arms. Study dosages were 10 to 40 mg/d. Dosages of placebo and fluoxetine hydrochloride began at 10 mg/d; the dosages were 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 the **Table**.

For venlafaxine IR, there were 11 adult studies with 2421 patients and 10 634 measurements; for venlafaxine ER, there were 10 adult studies with 2461 patients and 12 481 measurements. The majority of the studies were outpatient (venlafaxine IR, 2 inpatient and 9 outpatient; venlafaxine ER, 10 outpatient). Dosages were in the range of 25 to 375 mg/d (modal range, 75-150 mg/d). The Table provides a summary of the trials.

### STATISTICAL ANALYSIS

Analyses were conducted using SuperMix software.<sup>12</sup> Data were analyzed using a 3-level mixed-effects linear regression model.<sup>10</sup> 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 geriatric patients. Youths were not included in the combined analysis because they were assessed with the CDRS-R. 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 the treatment effect was tested by including a random treatment × time interaction. Time was the number of days from treatment initiation. The primary effect of interest was the 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 of the difference in HAM-D or CDRS-R scores at 6 weeks between treated and control groups. As a sensitivity analysis, we analyzed scheduled weekly visit data (ie, intended week of the visit instead of the actual day of the visit). We evaluated models with nonlinear time trends; however, the linear model provided the best overall fit to the data through 6 weeks. To test for the treatment effect's dependence on initial severity, we dichotomized baseline severity (HAM-D score<sup>13</sup> >20; CDRS-R score<sup>14</sup> >60) at the subject level and included interactions with treatment and time effects, with the estimate of interest being the 3-way severity × treatment × time interaction. The data were also analyzed using a continuous baseline severity score. These analyses included all 2- and 3-way interactions to determine whether baseline severity moderated the relationship between treatment and response. All analyses were adjusted for main effects of age and

**Table. Summary of Studies**

Study	Study Identification No.	Subjects, No.	No. (%)			Age, Mean (SD)	Dosage, mg/d	
			Treatment	Placebo	Female			
Adult fluoxetine study by Eli Lilly and Co	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	
	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	
	Geriatric fluoxetine study by Eli Lilly and Co	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	
Youth fluoxetine study by Eli Lilly and Co	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		221	109 (49.3)	112 (50.7)	118 (53.4)	14.6 (1.6)	10, 20, 40	
Adult venlafaxine ER study by Wyeth	014	202	100 (49.5)	102 (50.5)	121 (59.9)	40.3 (11.4)	37.5-225	
	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	
	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	
	Adult venlafaxine IR study by Wyeth	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		

Abbreviations: ER, extended release; IR, immediate release; TADS, Treatment for Adolescents With Depression Study.

sex, which decreases variability but does not determine whether 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 6 weeks for each patient and were compared using a mixed-effects logistic regression model adjusting for study. As a sensitivity analysis, the observed baseline and week 6 HAM-D or CDRS-R scores were used if available. Among all subjects, 52.2% 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 the HAM-D score at week 6 and remission was a HAM-D score lower than 8. For youths, response was a 50% reduction in the CDRS-R score at week 6 and remission was a CDRS-R score lower than 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 GERIATRIC PATIENTS

The estimated average rates of change over 6 weeks were -11.82 HAM-D units for drug vs -9.26 HAM-D units for placebo (MMLE=-2.55; SE=0.20;  $P < .001$ ), indicating 27.7% 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 the **Figure**. Variation in the treatment effect over studies was not statistically significant ( $SD=0.16$ ;  $P=.06$ ).

The estimated response rates were 58.4% for drug vs 39.9% for placebo (odds ratio [OR]=2.11; 95% CI, 1.93-2.31;  $P < .001$ ; NNT=5.41). Similar results were ob-

tained using available observed HAM-D scores (59.1% for drug vs 41.9% for placebo; OR=2.00; 95% CI, 1.83-2.19;  $P < .001$ ; NNT=5.82). Remission rates were 43.0% vs 29.3% for drug and placebo, respectively (OR=1.82; 95% CI, 1.66-2.00;  $P < .001$ ; NNT=7.30), and the sensitivity analysis revealed similar results (42.4% for drug vs 28.9% for placebo; OR=1.81; 95% CI, 1.65-1.99;  $P < .001$ ; NNT=7.39).

No effect of baseline severity on treatment efficacy was found for either the dichotomous ( $P = .27$ ) or continuous ( $P = .10$ ) baseline severity measures. For patients with low severity, the rates of change in symptoms over 6 weeks were  $-9.40$  HAM-D units for drug vs  $-7.20$  HAM-D units for placebo. For patients with high severity, the rates of change were  $-12.85$  for drug vs  $-10.07$  for placebo. The estimated differences were 2.20 HAM-D units (95% CI, 1.65-2.76) for low severity and 2.78 HAM-D units (95% CI, 2.26-3.29) for high severity. Estimated response rates for treated vs placebo-receiving patients were 54.8% vs 37.3% (difference of 17.5%), respectively, for low severity and 57.7% vs 40.5% (difference of 17.2%), respectively, for high severity. Estimated remission rates for treated vs placebo-receiving patients were 49.9% vs 36.6% (difference of 13.3%), respectively, for low severity and 37.8% vs 25.1% (difference of 12.7%), respectively, for high severity.

#### ADULT STUDIES OF FLUOXETINE

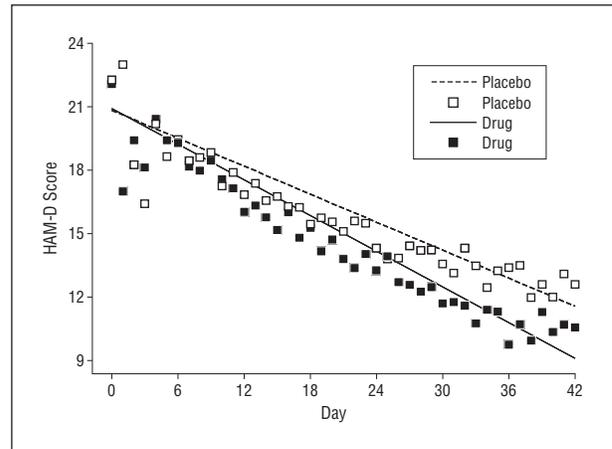
In adult studies of fluoxetine, the estimated average rates of change over 6 weeks were  $-10.12$  HAM-D units for fluoxetine and 7.52 HAM-D units for placebo (MMLE= $-2.60$ ; SE=0.34;  $P < .001$ ), indicating 34.6% greater improvement for fluoxetine.

The estimated response rates were 55.1% for fluoxetine vs 33.7% for placebo (OR=2.41; 95% CI, 1.93-3.01;  $P < .001$ ; NNT=4.69). Remission rates were 45.8% vs 30.2% for fluoxetine and placebo, respectively (OR=1.96; 95% CI, 1.66-2.31;  $P < .001$ ; NNT=6.40).

No effect of baseline severity on treatment efficacy was found ( $P = .14$ ). For patients with low severity, the rates of change in symptoms over 6 weeks were  $-7.98$  HAM-D units for fluoxetine vs  $-6.26$  HAM-D units for placebo. For patients with high severity, the rates of change were  $-11.72$  HAM-D units for fluoxetine vs  $-8.32$  HAM-D units for placebo. The estimated differences were 1.68 HAM-D units (95% CI, 0.77-2.59) for low severity and 3.40 HAM-D units (95% CI, 2.49-4.31) for high severity. Estimated response rates for treated vs placebo-receiving patients were 50.0% vs 36.4% (difference of 13.6%), respectively, for low severity and 56.2% vs 32.7% (difference of 23.5%), respectively, for high severity. Estimated remission rates for treated vs placebo-receiving patients were 51.7% vs 41.2% (difference of 10.5%), respectively, for low severity and 37.8% vs 22.1% (difference of 15.7%), respectively, for high severity. Neither response rates nor rates of improvement differed statistically between low- and high-severity groups.

#### ADULT STUDIES OF VENLAFAXINE ER

In adult studies of venlafaxine ER, the estimated average rates of change over 6 weeks were  $-12.39$  HAM-D



**Figure.** Observed vs estimated depression severity time trends for drug vs placebo in 37 adult and geriatric studies. HAM-D indicates Hamilton Depression Scale.

units for venlafaxine ER and  $-10.21$  HAM-D units for placebo (MMLE= $-2.18$ ; SE=0.38;  $P < .001$ ), indicating 21.4% greater improvement for venlafaxine ER.

The estimated response rates were 60.5% for venlafaxine ER vs 45.4% for placebo (OR=1.92; 95% CI, 1.61-2.29;  $P < .001$ ; NNT=6.63). Remission rates were 41.5% vs 29.5% for venlafaxine ER and placebo, respectively (OR=1.75; 95% CI, 1.46-2.09;  $P < .001$ ; NNT=8.31).

No effect of baseline severity on treatment efficacy was found ( $P = .94$ ). For patients with low severity, the rates of change were  $-9.58$  HAM-D units for venlafaxine ER vs  $-7.10$  HAM-D units for placebo. For patients with high severity, the rates of change were  $-12.81$  HAM-D units for venlafaxine ER vs  $-10.58$  HAM-D units for placebo. The estimated differences were 2.48 HAM-D units (95% CI, 0.91-4.05) for low severity and 2.23 HAM-D units (95% CI, 1.40-3.05) for high severity. Estimated response rates for treated vs placebo-receiving patients were 53.1% vs 36.6% (difference of 16.5%), respectively, for low severity and 59.9% vs 45.5% (difference of 14.4%), respectively, for high severity. Estimated remission rates for treated vs placebo-receiving patients were 52.1% vs 35.2% (difference of 16.9%), respectively, for low severity and 40.0% vs 28.6% (difference of 11.4%), respectively, for high severity.

#### ADULT STUDIES OF VENLAFAXINE IR

In adult studies of venlafaxine IR, the rates of change over 6 weeks were  $-14.32$  HAM-D units for venlafaxine IR and  $-10.71$  HAM-D units for placebo (MMLE= $-3.61$ ; SE=0.42;  $P < .001$ ), indicating 33.7% greater improvement for venlafaxine IR.

The estimated response rates were 67.2% for venlafaxine IR vs 45.2% for placebo (OR=3.16; 95% CI, 2.52-3.97;  $P < .001$ ; NNT=4.53). Remission rates were 47.1% vs 32.4% for venlafaxine IR and placebo, respectively (OR=2.24; 95% CI, 1.82-2.75;  $P < .001$ ; NNT=6.83).

No effect of baseline severity on treatment efficacy was found ( $P = .29$ ). For patients with low severity, the rates of change were  $-10.08$  HAM-D units for venlafaxine IR vs  $-7.69$  HAM-D units for placebo. For patients with high se-

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

## EFFECT OF AGE GROUP ON TREATMENT RESPONSE

### Geriatric Studies of Fluoxetine

In geriatric studies of fluoxetine, the estimated average rates of change over 6 weeks were -7.48 HAM-D units for placebo and -8.86 HAM-D units for fluoxetine (MMLE = -1.39; SE = 0.50;  $P = .009$ ), indicating 18.5% greater improvement for fluoxetine. However, response and remission rates were not significantly different between treated and control conditions.

The estimated response rates were 37.3% for fluoxetine vs 27.4% for placebo (OR = 1.42; 95% CI, 0.92-2.18;  $P = .12$ ; NNT = 16.95). Remission rates were 26.5% vs 20.0% for fluoxetine and placebo, respectively (OR = 1.26; 95% CI, 0.78-2.03;  $P = .34$ ; NNT = 38.71).

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

### Youth Studies of Fluoxetine

In youth studies of fluoxetine, the estimated average rates of change over 6 weeks were -15.96 CDRS-R units for placebo and -20.58 CDRS-R units for fluoxetine (MMLE = -4.62; SE = 1.26;  $P < .001$ ), indicating 29.0% greater improvement for fluoxetine.

The estimated response rates were 29.8% for fluoxetine vs 5.7% for placebo (OR = 6.66; 95% CI, 3.07-14.48;  $P < .001$ ; NNT = 4.16). Remission rates were 46.6% vs 16.5% for fluoxetine and placebo, respectively (OR = 4.23; 95% CI, 2.64-6.77;  $P < .001$ ; NNT = 3.33). The finding of higher remission rates 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 ( $P = .90$ ). For patients with low severity, the rates of change in symptoms were -17.60 CDRS-R units for fluoxetine vs -12.56 CDRS-R units for placebo. For patients with high severity, the rates of change were -28.86 CDRS-R units for fluoxetine vs -24.40 CDRS-R units for placebo. The estimated differences were 5.04 CDRS-R units (95% CI, 2.56 to 7.52) for low severity and 4.45 CDRS-R units (95% CI, -0.58 to 9.49) for high severity. Estimated response rates for treated vs placebo-receiving patients were 23.0% vs 3.2% (difference of 19.8%), respectively, for low severity and 40.2% vs 17.2% (difference of 23.0%), respectively, for high severity. Estimated remission rates for treated vs placebo-receiving patients were 54.1% vs 19.4% (difference of 34.7%), respectively, for low severity and 28.9% vs 7.5% (difference of 21.4%), respectively, for high severity.

## COMMENT

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 rates, response rates, and NNTs, the treatment effect was largest for youths followed by adults and more limited for geriatric patients. 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 control patients; this is similar to previous causal inference (growth mixture modeling) findings for fluoxetine and imipramine.<sup>15</sup> 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 control patients. Even stronger results were observed for children. In youth studies, 29.8% of treated children responded, whereas only 5.7% of children receiving placebo responded. Similarly, remission rates were 46.6% for treated children but only 16.5% for control patients. These rates translate into an additional child treated with fluoxetine responding and remitting for every 4 children treated. The higher rates of remission suggest that the remission criterion (CDRS-R score = 28) should be reevaluated.

By contrast, we found statistically significant (for HAM-D scores but not remission and response rates) but much less clinically significant effects for geriatric patients. Response rates were 37.3% for treated geriatric patients vs 27.4% for geriatric patients receiving placebo, translating to 1 additional patient responding to fluoxetine for every 17 patients treated. Remission rates were 26.5% for treated geriatric patients vs 20.0% for geriatric patients receiving placebo, which translates into 1 additional patient remitting with fluoxetine for every 39 patients treated. The efficacy of antidepressant treat-

ment 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 because both neuroendocrine challenge studies and receptor imaging studies report poorer antidepressant responses in depressed patients with more pronounced serotonin abnormality.<sup>16,17</sup> Serotonin function declines with age, potentially increasing the proportion of such patients in geriatric studies.

Venlafaxine produced results similar to those of fluoxetine, suggesting that these results are not specific to fluoxetine. Increased efficacy for the IR formulation compared with the ER formulation 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 end points 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, 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 by Frank et al,<sup>18</sup> have documented response rates for extended treatment with a single effective antidepressant. In that study, the remission rate was 82%, with 75% achieving remission by 140 days.<sup>17</sup> For fluoxetine, 23% of patients who were unimproved at 8 weeks showed full remission at 12 weeks.<sup>19</sup>

The findings of this study shed light on meta-analytic results that related the 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 end point, age group, or drug. Overall response rates were lower for geriatric patients than adults but did not vary by baseline severity. For children, response rates were lower overall compared with adults; however, here there were substantial differences between low and high baseline severity groups for both treated and control patients (ie, 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,<sup>1</sup> the use of average end points gleaned from studies with a variety of different approaches to handling missing data (eg, last observation carried forward 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 metaregression 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 article requires that all studies use a common end point. The use of different end points in different studies is exactly the type

of problem for which meta-analysis was designed 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.

This study has several limitations. First, we considered only 2 antidepressant medications, and other antidepressants may produce different effects. Indeed, fluoxetine is the only antidepressant approved for the treatment of childhood depression and is the only antidepressant that we studied in children. Second, there were only 4 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 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.<sup>20</sup> Third, a similar note of caution is in order for the results for the 4 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 the placebo benefit possibly degrades over time.<sup>19</sup> 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 that were designed to demonstrate efficacy and may have enrolled patients who may not have been representative of the patients seeking treatment for depression. However, 2 of the 4 youth studies were academic studies (Treatment for Adolescents With Depression Study and study X065). Because the largest effect of treatment was seen for youths and these studies are a mix of industry and academic studies, it seems unlikely that reliance on industry-sponsored studies produced biased results. Seventh, most of the studies had placebo lead-in periods that are designed to eliminate early placebo responders. However, an analysis<sup>21</sup> of 75 RCTs involving antidepressants and placebo from 1981 through 2000 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, these investigators 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 period on placebo response rates and 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 period.

To determine whether 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 MDD that met the following criteria: (1) not sponsored by a pharmaceutical company; (2) not associated with a specific medical illness (eg, following 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 HAM-D; and (6) had a minimum enrollment of 30 patients. Knowledge Finder (Aries System) 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 these 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 these criteria (n=1). Following these 2 manual reviews, reprints of candidate articles were obtained and reviewed for the 14 candidate articles. Two articles fulfilled the criteria.<sup>22,23</sup> The first study<sup>22</sup> was restricted exclusively to patients meeting the Columbia criteria for atypical depression (while meeting criteria for MDD). This study was partially funded by Eli Lilly and Co and their data were not available to us. The second study<sup>23</sup> 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 youths and smallest in geriatric patients (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 reanalyses 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 (eg, metaregression of study-level characteristics that exhibit interindividual variability, inconsistent and potentially biased handling of missing data). It 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.

**Submitted for Publication:** July 15, 2011; final revision received November 28, 2011; accepted December 1, 2011.  
**Published Online:** March 5, 2012. doi:10.1001/archgenpsychiatry.2011.2044

**Author Affiliations:** Departments of Medicine, Health Studies, and Psychiatry (Dr Gibbons) and Center for Health Statistics (Dr Gibbons, Hur, and Brown), University of Chicago, and Department of Psychiatry, Univer-

sity of Illinois at Chicago (Dr Davis), Chicago, and Center for Medication Safety, Pharmacy Benefit Management Services, Hines Veterans Affairs Hospital, Hines (Dr Hur); Prevention Science and Methodology Group, Center for Family Studies, Department of Epidemiology and Public Health, University of Miami, Miami, Florida (Dr Brown); and Department of Molecular Imaging and Neuropathology, New York State Psychiatric Institute and Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York (Dr Mann).  
**Correspondence:** Robert D. Gibbons, PhD, Center for Health Statistics and Departments of Medicine, Health Studies, and Psychiatry, University of Chicago, 5841 S Maryland Ave, MC 2007, Office W260, Chicago, IL 60637 (rdg@uchicago.edu).

**Author Contributions:** 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.

**Financial Disclosure:** Dr Gibbons has served as an expert witness for the US Department of Justice, Wyeth, and Pfizer Pharmaceuticals in cases related to antidepressants and anticonvulsants and suicide. Dr Brown directed a suicide prevention program at the University of South Florida that received funding from JDS Pharmaceuticals. Dr Mann has received research support from GlaxoSmithKline and Novartis.

**Funding/Support:** This work was supported by grants MH062185 (Dr Mann), R56 MH078580 (Drs Gibbons and Brown), MH8012201 (Drs Gibbons and Brown), and MH040859 (Dr Brown) from the National Institute of Mental Health and grant 1U18HS016973 from the Agency for Healthcare Research and Quality (Dr Gibbons).

**Additional Contributions:** Data were supplied by the National Institute of Mental Health (Treatment for Adolescents With Depression Study), Wyeth, and Eli Lilly and Co.

## 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(3):252-260.
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(2):e45.
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(1):47-53.
4. National Institute for Clinical Excellence. *Depression: Management of Depression in Primary and Secondary Care*. London, England: National Institute for Clinical Excellence; 2004.
5. Mathew SJ, Charney DS. Publication bias and the efficacy of antidepressants. *Am J Psychiatry*. 2009;166(2):140-145.
6. Melander H, Salomonson 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(9):623-627.
7. Brown CH, Wang W, Sandler I. Examining how context changes intervention impact: the use of effect sizes in multilevel mixture meta-analysis. *Child Dev Perspect*. 2008;2(3):198-305.
8. Lavori PW, Brown CH, Duan N, Gibbons RD, Greenhouse J. Missing data in longitudinal clinical trials, part A: design and conceptual issues. *Psychiatr Ann*. 2008;38(12):784-792.
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(2):227-246.
10. Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. New York, NY: John Wiley & Sons; 2006.

11. March J, Silva S, Petrycki S, Curry J, Wells K, Fairbank J, Burns B, Domino M, McNulty S, Vitiello B, Severe J; Treatment for Adolescents With Depression Study (TADS) 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(7):807-820.
12. Hedeker D, Gibbons RD, Du Toit SHC, Patterson D. *SuperMix: A Program for Mixed-Effects Regression Models*. Chicago, IL: Scientific Software International; 2008.
13. Elkin I, Gibbons RD, Shea MT, Sotsky SM, Watkins JT, Pilkonis PA, Hedeker D. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol*. 1995;63(5):841-847.
14. March JS. *Independent Evaluator Manual: Treatment for Adolescents With Depression Study (TADS), Final Version 4.1*. Durham, NC: Duke University Medical Center; 2005.
15. Muthén B, Brown HC. Estimating drug effects in the presence of placebo response: causal inference using growth mixture modeling. *Stat Med*. 2009;28(27):3363-3385.
16. Malone KM, Thase ME, Mieczkowski TA, Myers JE, Stull SD, Cooper TB, Mann JJ. Fenfluramine challenge test as a predictor of outcome in major depression. *Psychopharmacol Bull*. 1993;29(2):155-161.
17. Parsey RV, Olvet DM, Oquendo MA, Huang YY, Ogden RT, Mann JJ. Higher 5-HT<sub>1A</sub> receptor binding potential during a major depressive episode predicts poor treatment response: preliminary data from a naturalistic study. *Neuropsychopharmacology*. 2006;31(8):1745-1749.
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(12):1093-1099.
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? *Am J Psychiatry*. 2003;160(4):734-740.
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(15):1683-1696.
21. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA*. 2002;287(14):1840-1847.
22. McGrath PJ, Stewart JW, Janal MN, Petkova E, Quitkin FM, Klein DF. A placebo-controlled study of fluoxetine vs imipramine in the acute treatment of atypical depression. *Am J Psychiatry*. 2000;157(3):344-350.
23. Moreno RA, Teng CT, Almeida KM, Tavares Junior H. Hypericum perforatum vs fluoxetine in the treatment of mild to moderate depression: a randomized double-blind trial in a Brazilian sample. *Rev Bras Psiquiatr*. 2006;28(1):29-32.