Can People With Nonsevere Major Depression Benefit From Antidepressant Medication?

Jessica A. Stewart, MD; Deborah A. Deliyannides, MD; David J. Hellerstein, MD; Patrick J. McGrath, MD; and Jonathan W. Stewart, MD

ABSTRACT

Background: Several meta- or mega-analyses suggest antidepressant medications should be given only to severely depressed patients. In our experience, mild depression benefits from medication. We reanalyzed 1 clinic's randomized placebo-controlled antidepressant studies, limiting analyses to patients with major depressive disorder (MDD) without severe illness, to determine whether nonsevere depression responds to antidepressant medication.

Data Sources: Archives of the Depression Evaluation Service outpatient clinic of the New York State Psychiatric Institute were searched for randomized, placebo-controlled antidepressant studies that were conducted between 1977 and 2009 and included patients having MDD and pretreatment Hamilton Depression Rating Scale (HDRS) scores < 23.

Study Selection: Six placebo-controlled studies were found, including 8 active treatment arms and 1,440 patients. 825 patients were randomized and had MDD and an HDRS score < 23. *DSM-III, DSM-III-R,* or *DSM-IV* diagnostic criteria contemporary to each study were employed.

Data Extraction: Treatments were compared within study and via a patient-level meta-analysis using analysis of covariance (ANCOVA) of HDRS end point scores adjusted for pretreatment score. The number needed to treat (NNT) was calculated from remission rates (HDRS end point score ≤ 7), which were compared by χ^2. Effect sizes were calculated from change in HDRS scores. Secondary analyses investigated the effect of chronicity and atypical features on treatment response.

Data Synthesis: Three of 6 studies showed significant ($P < .001$) treatment effects by ANCOVA, and 4 of 6 studies showed significant ($P < .04$) differences in remission. The NNT ranged from 3 to 8. Effect sizes ranged from −0.04 to 0.8, with 4 of 8 greater than 0.4. The patient-level meta-analysis confirmed these results; neither chronicity nor atypical features significantly affected outcome. Secondary analyses utilizing global ratings and self-report mimicked the main findings.

Conclusions: Several studies demonstrated significant antidepressant efficacy for patients having nonsevere MDD. Efficacy was not trivial, as NNT ranged from 3 to 8, a range accepted by researchers as sufficiently robust to recommend treatment. These findings suggest mild-moderate MDD can benefit from antidepressants, contrary to findings by several other meta- or mega-analyses.

J Clin Psychiatry 2012;73(4):518–525

Submitted: *December 12, 2010; accepted May 9, 2011.*
Online ahead of print: *December 27, 2011 (doi:10.4088/JCP.10m06760).*
Corresponding author: *Jessica A. Stewart, MD, Beth Israel Deaconess Medical Center, Psychiatry Dept, Rabb-2, 330 Brookline Ave, Boston, MA 02215 (jstewar4@bidmc.harvard.edu).*

Marketing of antidepressants requires at least 2 "pivotal" randomized controlled trials (RCTs) demonstrating efficacy. Nevertheless, some doubt the utility of antidepressants, especially for treating patients with less than the most severe depression.[1–3] This negative view was highlighted by Fournier et al,[4] who reported that 2 antidepressants (imipramine and paroxetine) have a clinically significant effect only in patients having very severe depression (indicated by a Hamilton Depression Rating Scale [HDRS][5] score ≥ 23), while showing only negligible effects in those with mild, moderate, or severe depression.

Tedlow et al[6] previously showed baseline HDRS is inversely correlated with the probability of remission, without suggesting a cutoff below which efficacy may not be expected. Similar to Fournier et al,[4] Kirsch et al[7] concluded that statistical superiority of antidepressants relative to placebo is limited to the most severely depressed patients whose baseline HDRS score was ≥ 28. Fournier et al[4] limited their analyses to 6 of thousands of published studies, a limitation that raises questions of the representativeness and generalizability of their findings. Because Kirsch et al[7] used mean HDRS scores from 35 studies as their unit of analysis (meta-analysis), their proper inferences are to studies, not to individual patients or clinical practice. Khan et al,[8] analyzing a similar dataset, made recommendations only to the planning of future studies, carefully stating that inferences to clinical practice would be improper.

Despite the contrarian views of Fournier et al[4] and Kirsch et al,[7] some literature supports use of antidepressants in patients with nonsevere depression. For example, Silva de Lima et al,[9] reviewing 17 RCTs of dysthymia, reported efficacy of several classes of antidepressant medications and an overall number needed to treat (NNT) of 4. Similarly, reviewing the pharmacotherapy of nonmajor depression, Stewart et al[10] concluded that response rates to medications were similar to those seen in patients with major depression. Reanalyzing a study that included patients having major depression, Stewart et al[11] found both imipramine and phenelzine superior to placebo in patients having a mean ± SD HDRS score of 14 ± 4 (range, 7–33). Kocsis et al[12] reported that imipramine significantly outperformed placebo in patients with dysthymia, and imipramine was 4 times as likely to achieve remission (HDRS score ≤ 7) as placebo. Hellerstein et al[13] similarly reported patients having dysthymia achieved superior benefit if treated with fluoxetine relative to placebo. Khan et al[14] found a medium effect size in patients with an HDRS score < 23, but diagnostic criteria are not stated.

These studies demonstrating antidepressant efficacy in more mildly depressed patients were mostly conducted in

patients who did not have major depression. Yet questions about antidepressant efficacy are based on studies of major depression. We are not aware of studies limiting analysis to major depression with an HDRS score < 23. We therefore reanalyzed several RCTs, limiting the reanalyses to patients having both major depression and HDRS score < 23 to evaluate antidepressant response in patients with mild and moderate major depression. We hypothesized that such patients respond to antidepressant medication.

Clinical Points

- Antidepressants are effective for treating patients having nonsevere major depression.
- The number needed to treat for antidepressants in patients with nonsevere major depression is robust.

DATA SOURCES AND STUDY SELECTION

All studies conducted between 1977 and 2009 at the Depression Evaluation Service, an outpatient research clinic of the New York State Psychiatric Institute, were examined for inclusion. Inclusion criteria were studies that (1) included patients having major depression; (2) had no lower limit to entry HDRS score, or lower limit < 20; (3) randomized to at least 1 antidepressant versus placebo; and (4) included a database with at least prerandomization scores and end point scores on either HDRS-17 or HDRS-21.

Patients without major depression or with an HDRS score ≥ 23 were removed, leaving databases that included only patients having major depression and baseline HDRS-17 or HDRS-21 scores < 23. Because the HDRS-21 includes all items of the HDRS-17, patients having an HDRS-21 score < 23 must also have an HDRS-17 score < 23, and an HDRS-21 score ≤ 7 necessarily means HDRS-17 is also ≤ 7. That is, some patients having HDRS-21 scores > 22 might not have been excluded had their HDRS-17 been available, and some patients may have met remission criteria by HDRS-17 who did not by HDRS-21. Numbers of remitters may be underestimated in studies using HDRS-21, but effect on rates of remission is indeterminable since both numerator and denominator were underestimated. Because the HDRS-21 includes all the HDRS-17 items, the HDRS-17 could potentially be recalculated; however, some older databases included only 21-item summary scores and not individual items, so HDRS-17 could not be calculated. Where individual items were available, HDRS-17 was calculated and used in analyses. An HDRS end point score ≤ 7 defined remission for both HDRS versions.

Prior to 1985, diagnostic interviews used the Schedule for Affective Disorders and Schizophrenia (SADS)[15]; thereafter, the latest available version of the Structured Clinical Interview for *DSM* Disorders (SCID)[16–18] was used. Prior to 1980, SADS information generated Research Diagnostic Criteria diagnoses,[19,20] later converted by algorithm[21] to *DSM-III*.[22] After 1980, the current *DSM* diagnoses were generated, first from the SADS, later from the SCID. All participating patients signed informed consent, and the studies were approved by the New York State Psychiatric Institute's Institutional Review Board. All studies excluded patients having current (at least within last 6 months) drug or alcohol abuse or dependence, history of psychosis or medical illness affecting the brain, current unstable medical illness, and past adequate treatment with or prior intolerance to study medication.

After their initial evaluation, all patients received single-blind placebo for 1 to 2 weeks. Those who remitted or were found not to meet entry criteria were not randomized. Those who remained depressed and eligible were randomly assigned to double-blind treatment with active medication or matching placebo for 6 (4 studies: references 20, 22–26; and J. W. Stewart, unpublished data, 1996) or 12 (2 studies: reference 27 and J. W. Stewart, unpublished data, 1993) weeks. Studies conducted prior to 1984 obtained the HDRS-21 only at randomization and at study end point and only had the 21-item total score entered into the database. Thus, HDRS end point scores are missing on some patients who dropped out, and individual items are not available from which to calculate their 17-item scores. Study 6[27] similarly recorded only 17-item scores, individual items being unavailable for recalculating 21-item totals. For subjects without HDRS-17 scores, we performed the patient-level meta-analysis, approximating their HDRS-17 scores by multiplying their HDRS-21 scores by 0.9064, the conversion factor determined from the 498 subjects having both scores.

DATA EXTRACTION

Each study was reanalyzed for efficacy in subjects having major depressive disorder (MDD) and HDRS scores < 23. Reanalyses were conducted twice for each study, once for differences between drug and placebo based on HDRS end point score, covarying for baseline score using analysis of covariance, and once categorically for remission (defined as HDRS end point score ≤ 7) using χ^2, corrected for continuity when $df = 1$. The number needed to treat was calculated by inverting differences in remission rates between drug and placebo and rounding up to the next higher integer; effect size was calculated from the difference in HDRS change divided by group standard deviation. All analyses were 2-tailed, with an $\alpha = .05$. These analyses were also repeated combining the entire group and combining all nonplacebo treatments into 1 "active" group, which was compared to all patients receiving placebo.

Secondary analyses investigated baseline differences among studies that might account for any findings, as well as illness differences that may affect treatment response,

 © COPYRIGHT 2011 PHYSICIANS POSTGRADUATE PRESS, INC.

Table 1. Study Characteristics

Characteristic	All Patients	Study 1[20]	Study 2[22]	Study 3[23–26]	Study 4[a]	Study 5[b]	Study 6[27]
N	1,440	104	138	597	345	69	187
Study dates	9/77–5/96	9/77–11/78	2/79–4/81	11/79–10/87	3/88–5/96	7/91–3/93	1/93–7/95
HDRS version		21-Item	21-Item	21-Item	17-Item	17-Item	17-Item
DSM version		*III*	*III*	*III*	*III-R*	*IV*	*IV*
Medication(s)		Desipramine	Mianserin	Imipramine, phenelzine	Imipramine	Imipramine	Imipramine, fluoxetine
Inclusion/exclusion criteria							
Maximum HDRS score		18	18	None	None	None	None
Minimum HDRS score		None	None	None	None	None	10
No current MDD included		Yes	Yes	Yes	Yes	Yes	No
Melancholia included		Yes	No	Yes	Yes	Yes	No
No atypical depression included[c]		Yes	Yes	Yes	Yes	Yes	No[d]
Not randomized, n	256	24	29	96	55	19	34
Randomization, n							
HDRS score > 22	33	0	0	20	9	4	0
HDRS score < 23 and no current MDD	326	29	48	162	79	8	0
HDRS score < 23 and current MDD	825	51	61	319	203	38	153

[a]J. W. Stewart, unpublished data, 1996.
[b]J. W. Stewart, unpublished data, 1993.
[c]No atypical depression means depression without atypical features.
[d]Patients having significant mood reactivity plus 1 associated atypical feature (from among hyperphagia, hypersomnia, leaden paralysis, and pathological rejection sensitivity).
Abbreviations: HDRS = Hamilton Depression Rating Scale, MDD = major depressive disorder.

Table 2. Demographic and Clinical Characteristics of Patients Having HDRS Score < 23 and MDD

Characteristic	All Patients	Study 1[20]	Study 2[22]	Study 3[23–26]	Study 4[a]	Study 5[b]	Study 6[27]	Statistic
N	825	51	61	319	203	38	153	
Female, % (n/n)	63 (521/825)	53 (27/51)	44 (27/61)	65 (206/319)	67 (136/203)	47 (18/38)	64 (98/153)	$\chi^2_5 = 16.71$, $P < .01$
Age, mean ± SD, y	39 ± 11	41 ± 12	40 ± 11	38 ± 11	39 ± 11	40 ± 11	41 ± 10	$F_{5,819} = 3.11$, $P < .01$
Married, % (n/n)	29 (207/720)	41 (9/22)	21 (4/19)	29 (87/302)	28 (53/189)	47 (18/38)	24 (36/150)	$\chi^2_5 = 10.27$, $P < .1$
Caucasian, % (n/n)	89 (727/813)	78 (40/51)	92 (45/49)	92 (277/302)	90 (182/203)	89 (34/38)	88 (134/153)	$\chi^2_5 = 8.48$, NS
Employed, % (n/n)	51 (323/630)	83 (15/18)	67 (6/9)	40 (120/298)	61 (81/133)	50 (19/38)	63 (94/148)	$\chi^2_5 = 32.93$, $P < .001$
Years of education, mean ± SD	15 ± 3	14 ± 2	15 ± 3	14 ± 3	15 ± 3	14 ± 3	15 ± 3	$F_{5,616} = 2.98$, $P < .02$
Entry HDRS-21 score, mean ± SD	16 ± 4	16 ± 3	15 ± 3	16 ± 4	17 ± 4	18 ± 5	NA	$F_{4,667} = 4.60$, $P < .002$
Entry HDRS-17 score, mean ± SD	15 ± 4	NA	NA	14 ± 3	15 ± 4	16 ± 5	16 ± 4	$F_{3,709} = 10.13$, $P < .001$

[a]J. W. Stewart, unpublished data, 1996.
[b]J. W. Stewart, unpublished data, 1993.
Abbreviations: HDRS = Hamilton Depression Rating Scale, HDRS-17 = 17-item HDRS, HDRS-21 = 21-item HDRS, MDD = current major depressive disorder, NA = not available.

such as chronicity and atypical features. Further analyses determined whether measures besides the HDRS would corroborate the primary results.

DATA SYNTHESIS

Thirty-nine potential treatment studies were found. Of these, 20 did not include a group randomized to placebo, 3 excluded patients with major depression, 4 were discontinuation studies, and 6 were pharmaceutical company studies from which data were not released to the investigators. Six placebo-controlled studies of antidepressant medications were found (references 20, 22–27; J. W. Stewart, unpublished data, 1993; and J. W. Stewart, unpublished data, 1996) that included at least some patients with major depression who had pretreatment HDRS-17 or HDRS-21 scores < 23 and some patients randomly assigned to placebo. Diagnostic criteria used in the 6 studies to make a diagnosis of MDD were *DSM-III* (3 studies), *DSM III-R*[28] (1 study), and *DSM-IV*[29] (2 studies). Table 1 shows the 6 studies, including medication, relevant inclusion/exclusion criteria, number excluded, study duration, HDRS version (ie, 17-item or 21-item), and analyzed sample size (ie, number having major depression and HDRS score < 23). These studies included 1,440 depressed patients of whom 1,184 were randomized. Of the randomized patients, 359 were excluded from analyses because they had an HDRS score ≥ 23 or they did not meet criteria for MDD or both.

Table 2 shows the demographic and clinical characteristics of the patients. Mean ± SD age was 39 ± 11 years; 62% were women. Most were Caucasian, some had college education, and most were not currently married. About half were

Table 3. Medication Dosages[a]

Medication	Study 1[20]	Study 2[22]	Study 3[23–26]	Study 4[b]	Study 5[c]	Study 6[27]
Placebo imipramine, pills/d			5.4 ± 1.2	5.7 ± 0.9	5.7 ± 2.2	5.5 ± 1.3
Placebo non-imipramine, pills/d	5.8 ± 0.8	4.7 ± 0.9	5.4 ± 1.2			2.8 ± 0.6
Desipramine, mg/d	258 ± 69					
Imipramine, mg/d			258 ± 53	239 ± 77	285 ± 62	207 ± 89
Phenelzine, mg/d			67 ± 18			
Mianserin, mg/d		123 ± 54				
Fluoxetine, mg/d						50 ± 16

[a]Values are shown as mean ± SD.
[b]J. W. Stewart, unpublished data, 1996.
[c]J. W. Stewart, unpublished data, 1993.

Table 4. Comparisons of Hamilton Depression Rating Scale (HDRS) Scores and Remission Rates: Each Antidepressant Versus Placebo Within Study

			HDRS Score, Mean ± SD[b]			Remitted, % (n/n)[c]			Effect
Study	Treatment	n[a]	Baseline	End Point	Statistic		Statistic	NNT[d]	Size
1. Stewart et al,[20] 1985	Placebo	25	14.9 ± 3.2	11.1 ± 5.8	$F_{1,42} = 0.99$, NS	28 (7/25)	$\chi^2_1 = 1.40$, NS	6	0.15
	Desipramine	20	13.8 ± 2.6	8.8 ± 5.3		45 (9/20)			
2. McGrath et al,[22] 1985	Placebo	28	13.4 ± 3.0	10.4 ± 5.0	$F_{1,58} = 0.57$, NS	24 (6/25)	$\chi^2_1 = 4.46$, $P < .04$	4	0.20
	Mianserin	33	13.9 ± 2.2	9.4 ± 6.3		54 (14/26)			
3. Liebowitz et al,[23] 1988[e]	Placebo	91	14.1 ± 3.2	11.6 ± 5.5	$F_{2,173} = 17.27$, $P < .001$	27 (21/77)	$\chi^2_2 = 23.48$, $P < .001$		
Quitkin et al,[24] 1988	Imipramine	95	14.5 ± 3.5	9.9 ± 6.3		50 (45/90)		5	0.35
Quitkin et al,[25] 1990	Phenelzine	91	14.6 ± 3.4	7.1 ± 5.0		64 (53/83)		3	0.83
Quitkin et al,[26] 1989									
4. J. W. Stewart, 1996[f]	Placebo	93	13.1 ± 4.1	10.7 ± 5.7	$F_{1,195} = 22.40$, $P < .001$	20 (19/93)	$\chi^2_1 = 14.29$, $P < .001$	4	0.57
	Imipramine	105	12.9 ± 3.7	8.2 ± 5.2		49 (51/105)			
5. J. W. Stewart, 1993[f]	Placebo	16	16.1 ± 3.7	11.4 ± 6.6	$F_{1,43} = 0.32$, NS	39 (6/16)	$\chi^2_1 = 0.48$, NS	6	−0.04
	Imipramine	22	13.0 ± 4.3	8.6 ± 6.9		55 (12/22)			
6. McGrath et al,[27] 2000[g]	Placebo	52	13.9 ± 3.6	10.6 ± 6.1	$F_{2,148} = 9.68$, $P < .001$	38 (20/52)	$\chi^2_2 = 9.75$, $P < .01$		
	Imipramine	51	13.4 ± 3.2	5.8 ± 4.9		63 (32/51)		5	0.72
	Fluoxetine	49	13.8 ± 4.0	7.7 ± 5.6		51 (25/49)		8	0.47

[a]Sample sizes do not always agree with those shown in Table 1 due to missing data.
[b]HDRS-21 is shown for studies 1–3, HDRS-17 for studies 4–6.
[c]Percentage of remission is the percentage of those having an HDRS end point score ≤ 7.
[d]Number needed to treat = 1/percentage of patients remitted on drug − percentage of patients remitted on placebo.
[e]Imipramine versus placebo: $F_{1,173} = 9.56$, $P < .001$ (comparing HDRS); $\chi^2_1 = 6.53$, $P < .02$ (comparing percentage of remission). Phenelzine versus placebo: $F_{1,168} = 25.26$, $P < .001$(comparing HDRS); $\chi^2_1 = 19.72$, $P < .001$ (comparing percentage of remission).
[f]Unpublished data.
[g]Imipramine versus placebo: $F_{1,100} = 18.98$, $P < .001$ (comparing HDRS); $\chi^2_1 = 8.16$, $P < .003$ (comparing percentage of remission). Fluoxetine versus placebo: $F_{1,98} = 6.88$, $P < .02$ (comparing HDRS); $\chi^2_1 = 3.50$, $P < .1$ (comparing percentage of remission).
Abbreviations: HDRS = Hamilton Depression Rating Scale, NNT = number needed to treat, NS = not significant.

currently employed. Table 3 shows the mean medication dosage. As Table 2 shows, a number of patient characteristics varied across studies. Because some patients were missing documentation of varying amounts of these data, too few patients had all the data to include all variables in all analyses. Instead, separate covariate analyses were run for each variable. The analyses shown in Tables 4–7 do not include these characteristics as covariates because none demonstrated significant interactions with treatment on outcome in the individual covariate analyses. Neither chronicity nor presence or absence of atypical features significantly affected outcome ($P > .05$).

Table 4 shows within-study comparisons of remission rates of patients treated with antidepressants versus placebo. Three studies demonstrated significant end point drug-placebo differences in final HDRS scores, while 4 studies showed significant drug-placebo differences in remission rates as defined by HDRS end point score ≤ 7. The NNT ranged from 3 to 8, while effect sizes varied from −0.04 to 0.8, with 4 of 8 treatment arms having effect sizes of at least 0.4.

Table 5 shows secondary analyses of the doctor-rated Clinical Global Impressions-Improvement scale score (CGI-I)[30] and several self-rated measures, including a patient-rated version of the CGI-I, the Patient Global Impression of Improvement[31] (1 study); Beck Depression Inventory (BDI)[32] (2 studies); and the 90-item Symptom Checklist (SCL-90)[33] (4 studies). Nine of 13 drug-placebo comparisons were statistically significant.

Table 6 combines all patients from the 6 studies, analyzing for differences between active medication and placebo across various definitions of response. This is a patient-level meta-analysis, also known as a mega-analysis. All response definitions yielded significant results favoring active medication. The NNT ranged from 3 to 8.

Table 7 shows mean HDRS scores, SCL-90 summary score, SCL-90 depression subscale score, and BDI in a patient-level meta-analysis. Each measure showed

© Copyright 2011 Physicians Postgraduate Press, Inc.

Table 5. Comparisons of Global and Patient Ratings Antidepressant Versus Placebo Within Study

			Responders				SCL-90			BDI		
Study	Treatment	n	CGI, % (n/n)[a]	Statistic	PGI, % (n/n)[b]	Statistic	Pretreatment Score, Mean ± SD	Posttreatment Score, Mean ± SD	Statistic	Pretreatment Score, Mean ± SD	Posttreatment Score, Mean ± SD	Statistic
1. Stewart et al,[19] 1985	Placebo	26	31 (8/26)	$\chi^2_1 = 4.35$, $P < .04$	NA		NA	NA			NA	
	Desipramine	21	57 (12/21)									
2. McGrath et al,[21] 1985	Placebo	28	14 (4/28)	$\chi^2_1 = 5.49$, $P < .02$	NA		19.2 ± 4.0	16.3 ± 4.4	$F_{1,29} = 0.27$, NS		NA	
	Mianserin	33	45 (15/33)				19.4 ± 3.4	15.7 ± 3.8				
3. Liebowitz et al,[22] 1988[c]	Placebo	98	24 (23/94)	$\chi^2_2 = 51.44$, $P < .001$	NA		21.0 ± 5.9	20.2 ± 6.1	$F_{2,227} = 5.18$, $P < .01$		NA	
Quitkin et al,[23] 1988	Imipramine	96	47 (45/96)				20.7 ± 5.5	17.6 ± 6.0				
Quitkin et al,[24] 1990												
Quitkin et al,[25] 1989	Phenelzine	99	77 (72/94)				19.9 ± 5.1	14.2 ± 4.6				
4. J. W. Stewart, 1996[d]	Placebo	90	20 (17/83)	$\chi^2_1 = 12.72$, $P < .001$	NA		19.9 ± 4.7	17.9 ± 5.4	$F_{1,132} = 16.23$, $P < .001$	23.3 ± 7.6	19.2 ± 9.0	$F_{1,58} = 13.54$, $P < .001$
	Imipramine	101	47 (46/98)				18.9 ± 5.4	15.2 ± 4.5		23.1 ± 7.9	13.6 ± 9.7	
5. J. W. Stewart, 1993[d]	Placebo	16	31 (5/16)	$\chi^2_1 = 1.20$, NS	NA		22.9 ± 6.1	17.7 ± 6.1	$F_{1,32} = 1.44$, NS	22.8 ± 8.4	14.9 ± 8.6	$F_{1,34} = 0.59$, NS
	Imipramine	22	55 (12/22)				22.0 ± 5.7	15.3 ± 6.6		21.3 ± 8.4	11.8 ± 11.2	
6. McGrath et al,[26] 2000[e]	Placebo	52	25 (13/52)	$\chi^2_2 = 16.48$, $P < .001$	26 (13/50)	$\chi^2_2 = 15.9$, $P < .001$		NA			NA	
	Imipramine	52	62 (32/52)		63 (31/49)							
	Fluoxetine	49	57 (28/49)		57 (27/47)							

[a]A score of 1 ("very much improved") or 2 ("much improved") equals response.
[b]A score of 1 ("very much improved") or 2 ("much improved") equals response.
[c]In study 3, pairwise comparisons were used to analyze the percentage of responders according to the CGI (imipramine vs placebo: $\chi^2_1 = 9.42$, $P < .003$; phenelzine vs placebo: $\chi^2_1 = 49.03$, $P < .001$) and according to the SCL-90 (imipramine vs placebo: $F_{1,150} = 5.19$, $P < .003$; phenelzine vs placebo: $F_{1,150} = 9.64$, $df = 1$, $P < .003$).
[d]Unpublished data.
[e]In study 6, pairwise comparisons were used to analyze the percentage of responders according to the CGI (imipramine vs placebo: $\chi^2_1 = 12.69$, $P < .001$; fluoxetine vs placebo: $\chi^2_1 = 9.52$, $P < .003$) and according to the PGI (imipramine vs placebo: $\chi^2_1 = 12.45$, $P < .001$; fluoxetine vs placebo: $\chi^2_1 = 8.63$, $P < .004$).
Abbreviations: BDI = Beck Depression Inventory, CGI-I = Clinical Global Impressions-Improvement scale), NA = not available, NS = nonsignificant, PGI-I = Patient Global Impression of Improvement, SCL-90 = 90-item Symptom Checklist.

significant treatment differences in posttreatment scores covaried for pretreatment scores. Effect sizes based on these measures ranged from 0.46 to 0.66.

DISCUSSION

These data are most consistent with nonsevere major depression responding to a variety of antidepressant medications, including a tricyclic antidepressant (potent norepinephrine reuptake inhibitors), a monoamine oxidase inhibitor, a selective serotonin reuptake inhibitor, and a drug not marketed in the United States (mianserin, an inhibitor of multiple monoamine receptors[34]); each drug has an entirely different presumed mechanism of action. In many cases, effect sizes were moderate and the NNT was sufficiently low to make these reasonable treatment choices for such patients. These results conflict with suggestions that patients having nonsevere major depression are too unlikely to benefit from antidepressants for clinicians to consider the medications in their treatment planning. Why might our findings be so different from those of Fournier et al[4] and Kirsch et al,[7] and from which findings should clinicians extrapolate in their clinical practices?

The Fournier et al study[4] was a patient-level mega-analysis of 718 patients treated in 6 studies of 2 medications (paroxetine and imipramine) versus placebo. Fournier et al[4] question whether it is appropriate to offer antidepressant medications to patients having less than the most severe depression because they found the NNT was 11 for moderate depression and 16 for mild depression. However, these NNTs are based on 6 studies from among thousands of antidepressant RCTs. As they found 23 studies that met their inclusion criteria, it is unclear how representative the 6 studies were, even of these 23. Other problems with the Fournier et al[4] methodology include the inclusion of studies using low doses of medications, the exclusion of placebo run-in studies while including other delayed-treatment studies, and addressing only 2 of over 25 antidepressant medications and at least 8 different presumed mechanisms. While issues such as those raised by Fournier et al[4] are important, it also seems reasonable to wonder whether clinicians should change their practices based on conclusions made from possibly unrepresentative studies.

Reanalyzing 35 RCTs of new-generation antidepressants gleaned from US Food and Drug Administration (FDA) registration data, Kirsch et al[7] also concluded that only the most severely depressed patients benefit from antidepressant medications based on mean improvement in

Table 6. Mega-Analysis of Response: Active Medications (combined) Versus Placebo According to Various Definitions of Response

Response Definition	Placebo Responders, % (n/n)	Active Medication Responders, % (n/n)	χ^2	*P*	NNT
HDRS-17 end point score ≤ 7	29 (67/235)	57 (227/399)	46.77	< .001	4
HDRS-17 end point score ≤ 7 and change > 50%	20 (46/235)	49 (194/399)	51.82	< .001	4
HDRS-21 end point score ≤ 7	24 (58/238)	50 (180/363)	41.44	< .001	4
HDRS-21 end point score ≤ 7 and change > 50%	19 (46/248)	42 (150/361)	34.59	< .001	5
HDRS-17 end point score ≤ 7[a]	26 (80/308)	53 (250/473)	55.51	< .001	4
HDRS-17 end point score ≤ 7 and change > 50%[a]	17 (59/338)	44 (206/473)	48.44	< .001	4
BDI end point score < 10	18 (9/50)	43 (24/56)	6.50	< .02	4
BDI end point score < 10 and change > 50%	15 (7/48)	27 (28/103)	6.07	< .02	8
CGI-I end point score = 1 or 2	23 (70/309)	56 (266/477)	82.66	< .001	4
PGI end point score = 1 or 2	26 (13/50)	60 (58/97)	13.77	< .001	3

[a]All patients for whom an HDRS was available; when the individual items were available, the sum of the first 17 items was used; when only a 21-item summary score was available without individual items, the 21-item score was multiplied by 0.9 (the ratio of the 17-item score to the 21-item score in patients having both available). Note that numbers for HDRS-17 and HDRS-21 do not add up to HDRS-17 because some patients had 21 individual items so both the HDRS-17 and the HDRS-21 could be utilized.

Abbreviations: BDI = Beck Depression Inventory, CGI-I = Clinical Global Impressions-Improvement scale, HDRS = Hamilton Depression Rating Scale, NNT = number needed to treat, PGI = Patient Global Impression.

Table 7. Mega-Analysis of Comparing End Scores of All Patients on Active Medication Versus Placebo Covarying for Baseline Scores

Rating Scale Score[a]	Placebo	Antidepressant	Statistic	Effect Size
HDRS-17				
Pretreatment	13.74 ± 3.6	13.5 ± 3.6	$F_{1,629} = 56.32$, $P < .001$	0.52
Posttreatment	11.2 ± 5.6	7.7 ± 5.6		
HDRS-21				
Pretreatment	14.6 ± 3.6	14.3 ± 3.6	$F_{1,608} = 38.86$, $P < .001$	0.46
Posttreatment	11.8 ± 5.7	8.7 ± 6.0		
SCL-90 Summary				
Pretreatment	20.4 ± 5.5	20.1 ± 5.4	$F_{1,440} = 46.51$, $P < .001$	0.57
Posttreatment	18.8 ± 6.0	15.9 ± 5.5		
SCL-90 Depression subscale				
Pretreatment	3.2 ± 0.8	3.1 ± 0.8	$F_{1,411} = 60.52$, $P < .001$	0.66
Posttreatment	2.9 ± 1.0	2.2 ± 0.9		
BDI				
Pretreatment	23.0 ± 7.8	22.3 ± 8.1	$F_{1,94} = 10.11$, $P < .002$	0.63
Posttreatment	17.8 ± 9.1	12.7 ± 10.5		

[a]Values are shown as mean ± SD.

Abbreviations: BDI = Beck Depression Inventory, HDRS = Hamilton Depression Rating Scale, SCL-90 = 90-item Symptom Checklist.

HDRS scores, because significant effect sizes in improvement scores were found only when baseline HDRS severity exceeded 28. One problem with reaching such a conclusion is it is based on analyses of studies, not patients. Conclusions are best applied to planning future studies, not to clinical practice, as aptly noted by Khan et al[8] in reporting similar analyses of a larger FDA database. Other problems with the Kirsch et al[7] methods have been raised by Horder et al,[35] including misapplication and misinterpretation of statistical analyses.

There are several differences between our analyses and those of Fournier et al[4] and Kirsch et al.[7] First, our studies were all conducted at a single outpatient research clinic. Perhaps such a clinic attracts patients unrepresentative of patients seeking treatment in general practices. There is no way to know; yet the same sampling bias could be said of patients entering the studies used by Fournier et al[4] and Kirsch et al.[7] Furthermore, our use of a single site may address the ability to detect change in more mildly depressed patients. A single site with conscientious and uniformly trained raters improves reliability and decreases variance, thereby allowing detection of smaller changes in HDRS. Second, we analyzed each study separately, while Fournier et al[4] used patient-level mega-analysis and Kirsch et al[7] used more classical, study-level meta-analysis. Finally, we included data from a drug not marketed in the United States and from a monoamine oxidase inhibitor, drugs not included in the Fournier et al[4] and Kirsch et al[7] reports. Because we found benefits for drugs having 4 different putative mechanisms, our results could not have been driven by 1 drug or mechanism. This suggests our findings might apply to antidepressants in general. Because only 49 patients were treated with a "modern" agent, it can be argued our results may not generalize to treatments commonly used today. This objection does not diminish our implication that antidepressants can be effective in mildly depressed patients. It remains to be determined whether our conclusion applies to drugs introduced more recently than the older ones most of our patients received.

Two differences between our studies and those of Fournier et al[4] and Kirsch et al[7] are the absence of a major depression requirement for study entry in 5 of our studies and a lower HDRS entry requirement in 1 study. These differences suggest that clinicians should have had no incentive to apply diagnostic criteria loosely or inflate entry HDRS scores. Nevertheless, how representative patients coming to a single academic research center are of general clinical practice can be questioned; similarly, issues of generalizability from our results can legitimately be raised. At a minimum, however, our results query whether conclusions that mild or moderately depressed patients should not be offered antidepressants are justified.

One issue neither Fournier et al[4] nor Kirsch et al[7] discuss is the ease with which it is possible to detect change in the most severely depressed versus the more mildly depressed subjects. That is, someone starting with an HDRS score of 30 may drop to a score of 15 and still be depressed, while the maximum someone starting at a score of 14 can drop is 14 points; that is, a smaller maximal decrease than the still depressed 50% decrease of one with a starting

© Copyright 2011 Physicians Postgraduate Press, Inc.

score of 30. Because the HDRS has a floor, a more severely depressed person has more room to improve. And, because the HDRS can have a reliability of ± 3 points, it does not take much random fluctuation of symptoms for reliability plus random fluctuation of illness to produce an apparent 50% reduction or HDRS score ≤ 7 in a patient with an initially low HDRS patient score, while such "noise" will not result in such degrees of apparent change in patients having an initially high HDRS score. Thus, the question remains, how much do the findings of Fournier et al[4] and Kirsch et al[7] result from psychometric characteristics of the HDRS rather than from the inability of antidepressant medications to help patients who are less than the most severely depressed? That is, does failure to demonstrate a difference demonstrate a lack of difference versus a failure to detect a true difference?

One might wonder whether patients or clinicians can detect the changes captured by the HDRS—is the statistical difference an observable clinical change? One study[26] demonstrated a robust drug-placebo difference in response rates using the Patient Global Impression of Improvement. The finding that the physician-rated CGI-I also showed significant response differences in 5 of 6 studies, as did a conversion of the BDI into categorical response in 1 of 2 studies, suggests both physician and patient could notice differential change. The patient-rated SCL-90 also corroborated these conclusions in 2 of the 4 studies in which it was obtained (see Table 4).

Fournier et al[4] state that "there is little evidence to suggest that [antidepressants] produce specific pharmacologic benefit for the majority of patients with less severe acute depressions."(p52) They also state that the number needed to treat in severe depression is 4, which suggests that 75% of people with severe depressions also do not benefit. We agree with Fournier et al[4] that more predictably effective treatments are needed for all depressed subjects, including those with the most severe illness.

Drug names: desipramine (Norpramin and others), fluoxetine (Prozac and others), imipramine (Tofranil and others), paroxetine (Paxil, Pexeva, and others), phenelzine (Nardil and others).

Author affiliations: Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts (Dr J. A. Stewart), and New York State Psychiatric Institute and Columbia University, College of Physicians and Surgeons, New York, New York (Drs Deliyannides, Hellerstein, McGrath, and J. W. Stewart).

Potential conflicts of interest: **Dr J. A. Stewart** is an employee of Partners Healthcare. **Dr J. W. Stewart** has received grant/research support from National Institute of Mental Health (NIMH), Bristol-Myers Squibb, Boehringer-Ingelheim, Merck, Sanofi-Aventis, Roche, and Amylin; has served on speakers or advisory boards of Alkermes, Novartis, Sanofi-Aventis, Merck, and Boehringer-Ingelheim; and has received financial or material support for serving on Pfizer External Drug Monitoring Committee. **Drs Deliyannides**, **Hellerstein**, and **McGrath** report no financial or other potential conflicts of interest related to the subject of the article.

Funding/support: No funding was received for these reanalyses; one study on which these analyses are based was funded by NIMH, one by Organon, and one by Eli Lilly.

Previous presentations: Results of this study were presented in a poster session at the New Clinical Drug Evaluation Unit 50th Anniversary Meeting; June 14–17, 2010; Boca Raton, Florida.

Acknowledgment: We acknowledge the previous work of Depression Evaluation Service of the New York State Psychiatric Institute, whose data were analyzed for this study.

REFERENCES

1. Moncrieff J. Are antidepressants overrated? a review of methodological problems in antidepressant trials. *J Nerv Ment Dis*. 2001;189(5):288–295.
2. Healy D. The three faces of the antidepressants: a critical commentary on the clinical-economic context of diagnosis. *J Nerv Ment Dis*. 1999;187(3):174–180.
3. Fisher S, Greenberg RP. How sound is the double-blind design for evaluating psychotropic drugs? *J Nerv Ment Dis*. 1993;181(6):345–350.
4. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*. 2010;303(1):47–53.
5. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23(1):56–62.
6. Tedlow J, Fava M, Uebelacker L, et al. Outcome definitions and predictors in depression. *Psychother Psychosom*. 1998;67(4–5):266–270.
7. Kirsch I, Deacon BJ, Huedo-Medina TB, et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med*. 2008;5(2):e45.
8. Khan A, Leventhal RM, Khan SR, et al. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol*. 2002;22(1):40–45.
9. Silva de Lima M, Moncrieff J, Soares B. Drugs versus placebo for dysthymia (review). The Cochrane Collaboration. John Wiley & Sons, Hoboken, New Jersey, 2009 http://www.thecochranelibrary.com.
10. Stewart JW, Quitkin FM, Klein DF. The pharmacotherapy of minor depression. *Am J Psychother*. 1992;46(1):23–36.
11. Stewart JW, McGrath PJ, Quitkin FM. Can mildly depressed outpatients with atypical depression benefit from antidepressants? *Am J Psychiatry*. 1992;149(5):615–619.
12. Kocsis JH, Frances AJ, Voss C, et al. Imipramine treatment for chronic depression. *Arch Gen Psychiatry*. 1988;45(3):253–257.
13. Hellerstein DJ, Yanowitch P, Rosenthal J, et al. A randomized double-blind study of fluoxetine versus placebo in the treatment of dysthymia. *Am J Psychiatry*. 1993;150(8):1169–1175.
14. Khan A, Brodhead AE, Kolts RL, et al. Severity of depressive symptoms and response to antidepressants and placebo in antidepressant trials. *J Psychiatr Res*. 2005;39(2):145–150.
15. Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry*. 1978;35(7):837–844.
16. Williams JB, Spitzer RL, Gibbon M, et al. *Structured Clinical Interview for DSM-III*. New York, NY: Biometrics Research Department, New York State Psychiatric Institute; 1985.
17. Williams JB, Spitzer RL Gibbon M, et al. *Structured Clinical Interview for DSM-III-R*. New York, NY: Biometrics Research Department, New York State Psychiatric Institute; 1987.
18. First MB, Spitzer RL, Gibbon M, et al. *Structured Clinical Interview for DSM-IV Axis I Disorders–Patient Edition (SCID-I/P, Version 2.0)*. New York, NY: Biometrics Research Department, New York State Psychiatric Institute; 1996.
19. Spitzer RL, Endicott J, Robins E. Research diagnostic criteria: rationale and reliability. *Arch Gen Psychiatry*. 1978;35(6):773–782.
20. Stewart JW, McGrath PJ, Liebowitz MR, et al. Treatment outcome validation of *DSM-III* depressive subtypes: clinical usefulness in outpatients with mild to moderate depression. *Arch Gen Psychiatry*. 1985;42(12):1148–1153.
21. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition. Washington, DC: American Psychiatric Association; 1980.
22. McGrath PJ, Rabkin JG, Stewart JW, et al. Placebo-controlled study of mianserin in depressed outpatients. *Neuropsychobiology*. 1985;14(3):128–132.
23. Liebowitz MR, Quitkin FM, Stewart JW, et al. Antidepressant specificity in atypical depression. *Arch Gen Psychiatry*. 1988;45(2):129–137.
24. Quitkin FM, Stewart JW, McGrath PJ, et al. Phenelzine versus imipramine in the treatment of probable atypical depression: defining syndrome boundaries of selective MAOI responders. *Am J Psychiatry*. 1988;145(3):306–311.
25. Quitkin FM, McGrath PJ, Stewart JW, et al. Atypical depression, panic attacks, and response to imipramine and phenelzine: a replication. *Arch Gen Psychiatry*. 1990;47(10):935–941.
26. Quitkin FM, McGrath PJ, Stewart JW, et al. Phenelzine and imipramine in mood reactive depressives: further delineation of the syndrome of atypical depression. *Arch Gen Psychiatry*. 1989;46(9):787–793.

27. McGrath PJ, Stewart JW, Janal MN, et al. A placebo-controlled study of fluoxetine versus imipramine in the acute treatment of atypical depression. *Am J Psychiatry*. 2000;157(3):344–350.
28. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition, Revised. Washington, DC: American Psychiatric Association; 1987.
29. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Washington, DC: American Psychiatric Association; 1994.
30. Guy W. *ECDEU Assessment Manual of Psychopharmacology-Revised*. Rockville, MD: U.S. Department of Health, Education and Welfare; 1976.
31. Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry*. 2003;54(5):573–583.
32. Beck AT, Ward CH, Mendelson M, et al. An inventory for measuring depression. *Arch Gen Psychiatry*. 1961;4:561–571.
33. Derogatis LR. *SCL-9OR Administration, Scoring, and Procedures Manual*. Baltimore, MD: Clinical Psychometric Research; 1977.
34. Marshall RJ. The pharmacology of mianserin—an update. *Br J Clin Pharmacol*. 1983;15(suppl 2):263S–268S.
35. Horder J, Matthews P, Waldmann R. Placebo, Prozac and PLoS: significant lessons for psychopharmacology. *J Psychopharmacology*. 2010, Jun 22 [Epub ahead of print] jop.sagepub.com.

Editor's Note: We encourage authors to submit papers for consideration as a part of our Early Career Psychiatrists section. Please contact Marlene P. Freeman, MD, at mfreeman@psychiatrist.com.

© Copyright 2011 Physicians Postgraduate Press, Inc.