

Research Design Features and Patient Characteristics Associated With the Outcome of Antidepressant Clinical Trials

Arif Khan, M.D.

Russell L. Kolts, Ph.D.

Michael E. Thase, M.D.

K. Ranga Rama Krishnan, M.D.

Walter Brown, M.D.

Objective: The authors examined which, if any, research design features and patient characteristics would significantly differ between successful and unsuccessful antidepressant trials.

Method: Clinical trial data were reviewed for nine antidepressants approved by the Food and Drug Administration between 1985 and 2000. From the antidepressant research programs on these medications, 52 clinical trials were included in the study. The authors evaluated trial design features, patient characteristics, and difference in response between placebo and antidepressant.

Results: Nine trial design features and patient characteristics were present in the re-

search programs for all nine of the antidepressants. The severity of depressive symptoms before patient randomization, the dosing schedule (flexible versus fixed), the number of treatment arms, and the percentage of female patients were significantly associated with the difference in response to antidepressant and placebo. The duration of the antidepressant trial, number of patients per treatment arm, number of sites, and mean age of the patients were similar in successful trials (with a greater antidepressant-placebo difference) and less successful trials (with a smaller antidepressant-placebo difference).

Conclusions: These findings may help in the design of future antidepressant trials.

(*Am J Psychiatry* 2004; 161:2045–2049)

In over half of the recent clinical trials of antidepressants later approved by the Food and Drug Administration (FDA), the antidepressants failed to show an advantage over placebo (1). Part of the explanation might be in the increase of patients responding to placebo and antidepressants (to a lesser extent) in recent antidepressant clinical trials (2). However, it is unclear as to why the response to placebo and antidepressants is higher in recent trials than in earlier trials. One possible explanation that has been previously suggested is that the types of depressed patients participating in antidepressant clinical trials are changing (3).

Several other factors may contribute to the increase of patients responding to placebo and antidepressants in clinical trials. In previous reports, we have noted that factors such as the severity of mood symptoms at the time of entry into the trial (4) and the use of a flexible dose regimen, rather than a fixed dose regimen (5), may affect the outcomes of antidepressant clinical trials. A trial of shorter duration may produce a greater antidepressant-placebo difference than a trial of longer duration, as the response to placebo may be larger in longer trials (6). On the other hand, a shorter trial may not allow the full therapeutic effects of an antidepressant to occur.

Additionally, a greater number of treatment arms might increase the magnitude of the response to placebo (7). Some reports have suggested that female patients respond better to selective serotonin reuptake inhibitors (SSRIs)

than to tricyclic antidepressants, in part because of better tolerance (8, 9). However, other researchers have not been able to replicate this finding (10).

In order to explore whether these trial design factors and patient characteristics affected outcome among antidepressant clinical trials, we decided to explore the FDA summary basis of approval (SBA) reports, obtained by means of the Freedom of Information Act (11). We examined the trial design features and patient characteristics that were available in the FDA SBA reports on clinical trials of the nine antidepressants approved for sale in the United States between 1985 and 2000. We hypothesized that both trial design features and patient characteristics would differ significantly between successful antidepressant trials (those with a greater antidepressant-placebo difference) and less successful antidepressant trials (those with a smaller antidepressant-placebo difference).

Method

We obtained FDA clinical trial data (statistical and clinical reports) under the Freedom of Information Act (11) for nine antidepressants approved in the United States from Jan. 1, 1985, through Dec. 31, 2000: fluoxetine hydrochloride, sertraline hydrochloride, paroxetine hydrochloride, venlafaxine hydrochloride, nefazodone hydrochloride, mirtazapine, sustained-release bupropion hydrochloride, extended-release venlafaxine hydrochloride, and citalopram hydrobromide. The data were sent on microfiche or paper for a small fee in response to a specific request to the FDA,

TABLE 1. Characteristics of Clinical Trials for Nine Antidepressants Approved During 1985–2000, Based on FDA Summary Basis of Approval Reports

Investigational Antidepressant	Date of Approval	Number of Trials	Number of Investigational Arms	Number of Active Control Arms
Fluoxetine	Dec. 29, 1987	4	5	1
Sertraline	Dec. 6, 1991	3	5	2
Paroxetine	Dec. 29, 1992	11	11	6
Venlafaxine	Dec. 28, 1993	6	10	3
Nefazodone	Dec. 28, 1994	8	8	4
Mirtazapine	June 14, 1996	10	11	5
Sustained-release bupropion	Oct. 4, 1996	3	8	0
Extended-release venlafaxine	Oct. 20, 1997	3	3	2
Citalopram	July 17, 1998	4	8	0
Total		52	69	23

Freedom of Information Staff, 5600 Fishers Lane, HFI-35, Rockville, MD 20857. Some of the more recent clinical trial data were obtained over the Internet.

Of the research programs for the nine agents (fluoxetine, sertraline, paroxetine, venlafaxine, nefazodone, mirtazapine, sustained-release bupropion, extended-release venlafaxine, and citalopram), the FDA considered 56 clinical trials to be pivotal. Of these, we excluded four trials from our analysis. Three were excluded because of insufficient data, such as mean total scores on the Hamilton Depression Rating Scale (12), and one was excluded because it focused on relapse prevention rather than response to short-term treatment.

Among the remaining 52 trials, there were a total of 92 treatment arms, 69 investigational arms, and 23 active control arms (Table 1). We evaluated both trial design features and patient characteristics and found nine features that were present in all nine of the research programs: baseline depression severity, trial duration, flexible versus fixed doses, number of study sites, number of treatment arms, number of patients in each condition, patient age, percentage of female patients in the placebo group, and percentage of female patients in the antidepressant group. Features such as individual Hamilton depression scale scores, duration of depressive illness or episode, and past history were unavailable in the FDA clinical trial database. Mean values were calculated for each of the nine design features and patient characteristics.

For the purpose of analysis, each of the treatments containing a flexible dose was used as an independent unit. However, this formula was not followed for fixed-dose trials, as they had multiple treatment arms that all had set doses. For the trials that contained multiple treatment arms, mean scores were calculated across treatment arms, yielding a single score for each trial. Treatment arms with subtherapeutic doses (e.g., fluoxetine or paroxetine at 10 mg/day) were excluded from this analysis.

The difference between antidepressant and placebo in the mean change in the total score on the Hamilton depression scale was used to assess the successfulness of the antidepressant trial. We defined the antidepressant-placebo difference as follows: if the mean change (baseline through termination) in total Hamilton score was 12 in the antidepressant-treated group and the mean change was 8 in the placebo-treated group, then the antidepressant-placebo difference would be 4.

In our first analysis, we compared trial design features and patient characteristics using a median-split procedure to divide the trials into two groups on the basis of their antidepressant-placebo differences. Among the 52 trials, the mean antidepressant-placebo difference was 3.07 (range, –2.3 to 9.4). We divided

the trials into those among which the antidepressant-placebo difference was 3.07 or higher (N=26) and those among which the antidepressant-placebo difference was less than 3.07 (N=26). Thus, the 26 trials with the lower antidepressant-placebo difference scores (below the median score) made up the “less successful” group and were compared to the 26 trials designated “more successful” on the basis of higher antidepressant-placebo difference scores (above the median score).

To further characterize specific trial design features and patient characteristics of the antidepressant trials, we assessed factors by subdividing the trials into four equal quartiles on the basis of their mean antidepressant-placebo differences. We then conducted statistical analyses comparing the two most extreme groups: the group of 13 trials with the highest antidepressant-placebo difference scores was compared to the group of 13 trials with the lowest difference scores. The purpose of this analysis was to enable us to examine the design factors that differed between the most and least successful clinical trials.

We utilized t tests in cases where parametric statistics were appropriate to compare the design features of the “least successful” and “most successful” antidepressant clinical trials. We used Mann-Whitney U tests when the data were not appropriate for parametric analysis. In trials with missing data for select variables, we used pairwise deletion. Pairwise deletion allowed us to not use these trials for analyses that required the missing data but to include the trials in all other analyses. Finally, a correlational analysis was conducted to assess for the presence of any linear relationships between trial features and the degree of trial success as measured by the antidepressant-placebo difference.

Results

Of the 52 trials, 26 were grouped as “less successful” and 26 were grouped as “more successful.” The validity of this median-split procedure was supported by an expected significant difference between these groups in the antidepressant-placebo difference in the change in the total Hamilton depression scale score. Table 2 highlights the differences in design features and patient characteristics between the more successful trials and less successful trials. A higher percentage of the more successful trials used a flexible-dose design. Additionally, the more successful trials contained lower percentages of female patients in both the placebo and antidepressant groups. The more successful trials also included patients with higher Hamilton depression scores (more severe depression) at baseline. No differences were found with regard to trial length, number of sites, number of patients per treatment condition, or patient age.

In the second analysis, the 52 trials were divided into four groups by using a quartile split. The two most extreme groups (having the highest and lowest average antidepressant-placebo differences) were compared in regard to the nine common design features and patient characteristics (Table 3). As expected, the magnitude of the antidepressant-placebo difference scores on the Hamilton depression scale differed significantly between the two groups. As with our results based on a median split, the most successful trials were more likely to use a flexible dosing schedule, had lower percentages of female patients in both the placebo and antidepressant groups, and had higher Hamilton scores at baseline. The only difference between our results

TABLE 2. Features of 52 Antidepressant Trials With Less or More Success According to a Median Split of Antidepressant-Placebo Differences in Change in Hamilton Depression Rating Scale Scores^a

Design Feature or Patient Characteristic	Success of Trial, According to Median Split of Antidepressant-Placebo Differences in Change in Hamilton Score, With Last Observation Carried Forward				Significance		
	Less Successful (26 trials)		More Successful (26 trials)		t	df	p
	Mean	SD	Mean	SD			
Baseline score on Hamilton Depression Rating Scale	24.94	1.61	26.38	2.61	2.39	50	<0.05
Antidepressant-placebo difference in change in Hamilton depression score, with last observation carried forward	1.30	1.25	4.84	1.67	8.76	50	<0.001
Length of trial (weeks)	6.23	1.14	6.04	1.66	0.49	50	0.63
Number of sites	5.33	4.94	3.73	3.75	1.26	45	0.21
Number of treatment arms	3.19	0.98	2.77	0.65	1.83	50	0.07
Mean number of patients per condition	73.77	36.21	57.44	26.92	1.80	48	0.08
Mean age of patients (years)	43.11	7.81	41.98	6.68	0.52	43	0.61
	Mean	SD	Mean	SD	U	N	p
Percent of patients who were female							
Placebo group	61.93	9.81	54.36	11.98	165.5	45	<0.05
Antidepressant group	62.89	7.64	54.35	11.04	127.0	44	<0.01
	N	%	N	%	U	N	p
Trials using flexible doses	15	57.7	22	84.6	221.0	52	<0.01

^a Data on some variables were not available for all trials.

based on the median split and the quartile split was the finding based on the quartile split that the most successful trials used fewer treatment arms. Again, no difference was observed with regard to trial length, number of sites, number of patients per condition, or patient age.

Additionally, we examined the ranges of the data for these variables to observe whether the extent of the ranges may have influenced the results. The range of the mean baseline Hamilton depression score was 21.6 to 33.6. Trial length varied from 4 weeks to 12 weeks, and the number of treatment arms ranged from 2 to 5. The number of sites ranged from 1 to 18. The mean number of patients per condition ranged from 21 to 172, while the range of the mean patient age was 33.0 to 77.1 years.

We also conducted a correlational analysis to assess for the presence of linear relationships between trial features and the degree of response as measured by the difference between antidepressant and placebo in the change in the Hamilton depression scores, with the last observation carried forward (Table 4). A larger antidepressant-placebo difference was positively associated with a higher baseline Hamilton depression score and the use of flexible dosing schedules. Additionally, antidepressant-placebo difference was significantly negatively associated with the number of treatment arms and the percentages of female patients in both the placebo and antidepressant groups. No relationship was observed between outcome and trial length, number of sites, number of patients, or patient age.

Discussion

The aim of our study was to assess the existence of design features and patient characteristics in antidepressant

clinical trials that might be associated with clinical trial outcome. Our analysis suggests that greater severity of depressive symptoms before randomization, flexible dosing schedule (versus fixed doses), fewer treatment arms, and a lower percentage of female patients were significantly associated with successful outcome, as defined by the difference between antidepressant and placebo in the change in the total score on the Hamilton Depression Rating Scale.

It is not surprising that we found greater severity of depressive symptoms at baseline and flexible dosing to be associated with greater success in antidepressant trials. We reported such phenomena in our earlier analysis of the FDA SBA reports (1, 5), and our results support the previous finding (7) that a higher number of treatment arms is associated with a greater magnitude of response to placebo. This in turn is likely to reduce the chances of a successful antidepressant trial. However, it is not clear which antidepressant trial design features and patient characteristics mutually exist in the FDA SBA reports and the published reports that were previously reviewed (7).

Although studies have suggested (8, 9) that women and men may respond differently to antidepressants, we found an unexpected and paradoxical phenomenon. Among the FDA SBA reports, antidepressant trials with fewer women were more successful than trials with more women. Alternatively, antidepressant trials with more men were more successful than trials with fewer men. This implies that antidepressant-placebo differences were larger among men than among women.

However, we cannot adequately substantiate this finding as the FDA SBA reports did not report individual scores and did not present scores in relation to the sex of the participating patients. This phenomenon was in part due to

TABLE 3. Features of 26 Antidepressant Trials With the Least and Most Success According to a Quartile Split of Antidepressant-Placebo Differences in Hamilton Depression Rating Scale Scores^a

Design Feature or Patient Characteristic	Success of Trial, According to Quartile Split of Antidepressant-Placebo Differences in Change in Hamilton Score, With Last Observation Carried Forward				Significance		
	Least Successful (13 trials)		Most Successful (13 trials)				
	Mean	SD	Mean	SD	t	df	p
Baseline score on Hamilton Depression Rating Scale	24.26	1.42	26.92	2.40	3.45	24	<0.01
Antidepressant-placebo difference in change in Hamilton depression score, with last observation carried forward	0.48	1.23	6.17	1.23	11.79	24	<0.001
Length of trial (weeks)	6.23	1.17	6.08	2.10	0.23	23	0.82
Number of sites	5.82	6.05	3.77	4.40	0.96	22	0.35
Number of treatment arms	3.15	0.99	2.46	0.52	2.24	24	<0.05
Mean number of patients per condition	64.54	24.83	49.38	21.16	1.64	23	0.12
Mean age of patients (years)	44.29	10.70	41.14	5.77	0.93	23	0.36
	Mean	SD	Mean	SD	U	N	p
Percent of patients who were female							
Placebo group	66.85	8.41	54.04	12.04	27.0	24	<0.01
Antidepressant group	64.10	8.28	56.03	10.09	36.0	24	<0.05
	N	%	N	%	U	N	p
Trials using flexible doses	8	61.5	13	100.0	52.0	26	<0.05

^a Data on some variables were not available for all trials.

TABLE 4. Correlations Between Trial Features and Magnitude of Antidepressant-Placebo Difference in Change in Hamilton Depression Rating Scale Scores in 52 Antidepressant Trials

Design Feature or Patient Characteristic	Correlation With Antidepressant-Placebo Difference		
	r	N	p
Baseline score on Hamilton Depression Rating Scale	0.42	52	0.002
Dosing schedule (flexible versus fixed)	0.36	52	0.01
Trial length	-0.05	52	0.75
Number of treatment arms	-0.33	52	0.02
Number of sites	-0.08	47	0.61
Percent of patients in placebo group who were female	-0.36	45	0.02
Percent of patients in antidepressant group who were female	-0.31	44	0.05
Mean number of patients per condition	-0.15	50	0.29
Mean patient age	-0.06	45	0.69

the FDA's reluctance to include women of childbearing potential in the 1980s.

We were surprised to find that the duration of antidepressant trial, number of patients per treatment arm, and number of sites were not related to the outcome of the trial. Furthermore, the ages of the patients were similar in the successful and not so successful trials. However, the age distributions were similar among most trials, and thus we cannot comment on the potential effects of including either geriatric or pediatric populations. In short, we may have failed to detect the possible role of these research design features because of the limitations of the FDA SBA report data.

A number of design features, most notably dosing schedule and number of trial arms, were highly intercorrelated, making it difficult to assess the unique contribution of each

feature to trial outcome. Again, we were not able to assess many other possible antidepressant trial features and patient characteristics that may be associated with trial success as these were not available in the FDA SBA reports. These include the role of various rating scales, including modified versions of the Hamilton Depression Rating Scale, the Montgomery-Åsberg Depression Rating Scale, and other scales. It is possible that trial results may differ among various countries and cultures and also that individual patient characteristics may be different among various studies. Such features may include the frequency of melancholic depression, chronicity of depressive episodes or depressive illness, and history of resistance to antidepressant treatment. For example, Zimmerman et al. (13) elegantly showed that fewer than 30% of depressed patients seen in clinical practice can be included in antidepressant clinical trials. Thus, our findings are limited to clinical trial populations, rather than to all depressed patients.

In summary, we found that design features of antidepressant trials, such as severity of symptoms before randomization, use of flexible dosing of antidepressants, and fewer treatment arms, were observed significantly more frequently among successful trials. Additionally, successful trials contained a higher number of men than women. These findings may help in the design of future antidepressant trials.

Received Oct. 13, 2003; revision received Dec. 29, 2003; accepted March 10, 2004. From the Northwest Clinical Research Center; the Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, N.C.; the Department of Psychology, Eastern Washington University, Cheney, Wash.; the Department of Psychiatry, University of Pittsburgh School of Medicine; the Department of Psychiatry, Brown University Medical School, Providence, R.I.; and the Department of Psychiatry, Tufts University School of Medicine,

Boston. Address reprint requests to Dr. Khan, Northwest Clinical Research Center, Number 112, 1900 116th Avenue NE, Bellevue, WA 98004; akhan@nwcrc.net (e-mail).

The authors thank Amy Brodhead, M.S., for assistance with the manuscript.

References

1. Khan A, Khan S, Brown WA: Are placebo controls necessary to test new antidepressants and anxiolytics? *Int J Neuropsychopharmacol* 2002; 5:193–197
2. Walsh BT, Seidman SN, Sysko R, Gould M: Placebo response in studies of major depression. *JAMA* 2002; 287:1840–1847
3. Robinson DS, Rickels K: Concerns about clinical drug trials. *J Clin Psychopharmacol* 2000; 20:593–596
4. Khan A, Leventhal RM, Khan SR, Brown WA: Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol* 2001; 22:40–45
5. Khan A, Khan S, Walens G, Kolts R, Giller E: Frequency of positive studies among fixed and flexible dose antidepressant clinical trials: an analysis of the Food and Drug Administration summary basis of approval reports. *Neuropsychopharmacology* 2003; 28:552–557
6. Khan A, Warner H, Brown WA: Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: an analysis of the FDA database. *Arch Gen Psychiatry* 2000; 57:311–317
7. Zimmerman M, Posternak MA: Placebo response in antidepressant efficacy trials: relationship to number of active treatment groups, in 2003 Annual Meeting New Research Program and Abstracts. Arlington, Va, American Psychiatric Association, 2003, number 893
8. Kornstein SG, Schatzberg AF, Thase ME, Yonkers KA, McCullough JP, Keitner GI, Gelenberg AJ, Davis SM, Harrison W, Keller MB: Gender differences in treatment response to sertraline versus imipramine in chronic depression. *Am J Psychiatry* 2000; 157:1445–1452
9. Kornstein SG, Sloan DM, Thase ME: Gender-specific differences in depression and treatment response. *Psychopharmacol Bull* 2002; 36:99–112
10. Quitkin FM, Stewart JW, McGrath PJ, Taylor BP, Tisminetzky MS, Petkova E, Ma YG, Klein DF: Are there differences between women's and men's antidepressant responses? *Am J Psychiatry* 2002; 159:1848–1854
11. Freedom of Information Act. 5 Congress 552 (1994 and Supp II 1996). <http://www.usdoj.gov/04foia/>
12. Hamilton M: A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; 23:56–62
13. Zimmerman M, Mattia JJ, Posternak MA: Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *Am J Psychiatry* 2002; 159: 469–473