Can the Effects of Antidepressants Be Observed in the First Two Weeks of Treatment?

Frederic M. Quitkin, M.D., Patrick J. McGrath, M.D., Jonathan W. Stewart, M.D., Bonnie P. Taylor, M.A., and Donald F. Klein, M.D.

Recently the claim that there is a lag in the time of onset of antidepressant effects has been questioned. This issue rests on contrasting the time course of ultimate responders versus nonresponders on imipramine and amitriptyline. It is concluded that in 1 week on antidepressants "clinicians were capable of detecting changes in general states between the groups and the specific effects of depressed mood and anxiety and the physical expression of distress" (Katz et al. 1987). To examine this issue, we first used the design in which ultimate responders and ultimate nonresponders to antidepressants were compared at 1 and 2 weeks. Clearly there were statistically significant differences between

ultimate responders and nonresponders on drug. However, the same was true on placebo. When the ultimate responders on placebo were contrasted to the ultimate responders on drug at 1 and 2 weeks using the Clinical global Impression (CGI) scale and the Hamilton Depression Rating Scale (HDRS), there was no difference between drug and placebo. This was also true for a subgroup of patients who met the criteria for melancholia. We conclude that, if the effects of nonspecific improvement are partialed out, there is no evidence of a medication effect at 1 and 2 weeks.
[Neuropsychopharmacology 15:390–394, 1996]

Tricyclic antidepressants (TCA) and monoamine oxidase inhibitors (MAOI) were developed in the 1950s. With their introduction, pharmacological agents have become the cornerstone in the treatment of depressive disorders. Depending on the method of categorization, there are now four or five classes of antidepressant compounds. These include the TCAs, the MAOIs, the selective serotonin reuptake inhibitors (SSRIs), the tetracyclics, and the chemically unrelated compounds with antidepressant properties (bupropion, venlafaxine, nefazadone,

etc.) generally referred to as second-generation drugs. All of these drugs are quite effective, with an h (effect size index) in the moderate range (30%–40% better than placebo, $h = .5$–$.6$; Klein et al. 1980; Janicak et al. 1993).

In contrast to establishing the clear efficacy of these drugs, characterizing symptom profiles of patients responsive to different classes of drugs has been imprecise (Bielski and Friedel 1976; Joyce and Paykel 1989). Although 60% or 70% of patients respond to the first drug they receive, some require multiple trials (McGrath et al. 1993). If patient and physician are steadfast, eventually a beneficial class of antidepressants will be identified for most patients. A limitation in prescribing successive trials is the fact that each may take 4 to 6 weeks before a drug's lack of efficacy is apparent and a new one is prescribed. Not uncommonly, there may be a delay of several months before an effective drug is found, thus, the interest in identifying early signs of improvement or lack of it, so that an alternate drug can be prescribed. There also are important heuristic issues concerning the time of onset of specific drug action.

It has been suggested that available drugs have some effects that may be observed within 1 week. Winokur (1980) reviewed several studies that presented a mean improvement on the Hamilton Depression Rating Scale (HDRS) each week for drug-treated patients. He concluded that the usual response for patients treated with antidepressants appeared to be immediate, "at least immediate in terms of 1 week" and might be attributable to the natural course of the illness. Similar observations have been reported (Coryell et al. 1982; Katz et al. 1987; Casper et al. 1994; Stassen et al. 1994; Katz et al. 1994). A contrary position, according to which there is at least a delay of several weeks in the onset of true drug effect, has been suggested by others (Quitkin et al. 1987).

Katz et al. (1987) have been the strongest adherents to the view that specific changes by 1 week allow the clinician to distinguish responders versus nonresponders to antidepressants. This claim rests on contrasting the time course of ultimate responders versus nonresponders to imipramine and amitriptyline. In 1 week clinicians "were capable of detecting change in the general state between the groups in the specific effects of depressed mood and anxiety and the physical expression of distress" (Katz et al. 1987). They suggest that effect sizes are large, and the implication is that antecedent change could be used to detect ultimate responders.

Katz et al. (1987, 1994) estimated effect size at week 1 to vary from .24 to .9 depending on the measure. They used guidelines from Cohen (1988), which suggest that an effect size of .5 is medium and readily detectable to a clinical observer. However, their conclusions are open to question because, without a placebo control, they cannot estimate differential improvement attributable to nondrug, nonspecific effects (placebo effects and spontaneous remission). The potency of nonspecific effects in all forms of psychiatric and, indeed, all medical treatments has long been recognized (this is particularly true in disorders characterized by a phasic or fluctuating course). Until fairly recently, in fact, physicians had little else to work with in the treatment of any disorder.

The purpose of this report is to assess critically the claim of Katz et al. that contrasting ultimate responders versus ultimate nonresponders permits a clear observation of true drug effects by 1 week. In contrast to this earlier work, we can partial out the effects of nonspecific improvement because our studies included a placebo group.

Question 1. Would contrast of ultimate responders versus nonresponders within drug and placebo groups show a difference in outcome at weeks 1 and 2 on the Clinical Global Impression (CGI) scale? Would there also be a treatment effect?

Question 2. Would contrasting the HDRS for ultimate responders versus nonresponders in drug and placebo trials show a difference at week 2? Would there also be a treatment effect?

Question 3. What is the effect size if ultimate responders and nonresponders are contrasted on the proportion rated responders at week 1 and week 2 in drug and placebo trials? What is the effect size if the proportion of all patients (ultimate responders and nonresponders combined) rated responders at week 1 and week 2 on drug is contrasted with all patients on placebo?

METHODS

We examined all studies conducted during a 10-year period in the Depression Evaluation Service, an outpatient research clinic affiliated with Columbia University at the New York State Psychiatric Institute (Stewart et al. 1980, 1981; McGrath et al. 1984, 1985; Quitkin et al. 1984, 1989; Liebowitz et al. 1988). All studies included a single-blind, 7- to 10-day placebo run-in period, and a 6-week double-blind trial with a fixed flexible dose schedule. There were five drug-placebo contrasts and three drug–drug contrasts. The medications used included three MAOIs (phenelzine, tranylcypromine, and deprenyl), three TCAs (imipramine, desipramine, and amitriptyline and mianserin. At the end of the 7- to 10-day initial placebo run-in period (IPR), all patients' clinical status was reassessed, and change was measured on the CGI (a weekly overall rating of mood change from the treatment baseline). Ratings are made on a 7-point scale (1 = very much improved, 2 = much improved, 3 = minimally improved, 4–7 = unchanged or worse). Patients rated 1 or 2 (at least much better) after IPR did not enter the double-blind phase. This analysis included patients who completed 6 study weeks. Five hundred and ninety-three patients who were minimally improved, unimproved, or worse completed 6 weeks in a double-blind study. The HDRS (21-item) was completed at baseline, 4 and 6 weeks in all studies. In some studies, it was also performed at week 2.

A patient was considered a responder if he or she was rated at least "much improved" at week 6 in the physician-rated improvement scale of the CGI. These patients had a significant amelioration in depressed affect and a minimum of 75% reduction in intensity and frequency of depressed mood. All patients who received a 3 on the CGI were considered minimally improved. These patients were noted to have some improvement, but most of their depressive affect persisted.

RESULTS

Question 1

Would contrast of ultimate responders versus nonresponders in drug and placebo groups show a difference in outcome at weeks 1 and 2 on the CGI? Would there also be a treatment effect? To simulate the analysis by

Table 1. Contrasts of the Proportion of Responders at Weeks 1 and 2 for Ultimate Responders (week 6) and Nonresponders (ANOVA of proportions)

	Ultimate (week 6) Responders Improved	Ultimate (week 6) Nonresponders Improved
Week 1 analysis[a]		
Placebo	21% (12/58)	7% (10/143)
Drug	13% (32/252)	6% (9/140)
Week 2 analysis[b]		
Placebo	33% (19/58)	8% (12/143)
Drug	29% (74/252)	10% (14/140)

[a] X^2 for responder versus nonresponder (between columns ignoring rows): $X^2 = 10.05$, df $= 1$, $p = 0.001$; X^2 for treatment (between rows ignoring columns): $X^2 = 1.17$, df $= 1$, $p = 0.278$; X^2 interaction: $X^2 = 1.06$, df $= 1$, $p = 0.304$.
[b] X^2 for responder versus nonresponder (between columns ignoring rows): $X^2 = 38.625$, df $= 1$, $p = 0.000$; X^2 for treatment (between rows ignoring columns): $X^2 = .001$, df $= 1$, $p = .964$; X^2 interaction: $X^2 = .471$, df $= 1$, $p = 0.492$.

Katz et al., the clinical status of patients on drug and placebo who were and were not responders at 6 weeks as contrasted at week 1 (Table 1). For example, in the placebo group at week 1 there were 58 responders at week 6, and 12 (21%) were much improved at week 1. There were 145 nonresponders at week 6, and 10 (7%) were improved at week 1. An analysis of variance (ANOVA) of proportions was performed. There is a clear effect for responder versus nonresponder (between columns ignoring rows, $p = .001$), no treatment effect (between rows ignoring columns, $p =$ NS), and no significant interaction. The same analysis was performed at week 2 with virtually identical results (Table 1).

A similar analysis was done limiting the sample to patients meeting DSM-III or DSM-III-R criteria for melancholia. the outcome is similar to the analysis from the entire patient sample. There is a main effect of ultimate

responder status, no treatment effect, and no interaction (ANOVA proportions, responder effect $X^2 = 24.9$, df $= 1$, $p = .000$; treatment effect $X^2 = 1.7$, df, $p =$ NS) and the interaction was not significant.

Question 2

Would contrasting the HDRS for ultimate responders versus nonresponders in drug and placebo trials show a difference at week 2? Would there also be a treatment effect? A two-way ANOVA of contrasting the HDRS at 2 weeks for ultimate responders versus nonresponders in placebo and drug trials was done (Table 2). The covariate was the HDRS at baseline, and the dependent variable was week-2 HDRS scores. At 2 weeks, (drug and placebo) there are significant differences between ultimate responders and ultimate nonresponders. The main effect of ultimate responder versus ultimate non-

Table 2. Contrast of Week 2 HDRS Scores in Ultimate Responders and Nonresponders (two-way ANCOVA)

	HDRS Score at RZ (Mean ± SD)	HDRS Score at Week 2 (Mean ± SD)	F
Main effects			
Response			
IMI R	14.5 ± 4.1	9.7 ± 4.9	
IMI NR	15.1 ± 3.9	12.5 ± 4.2	
PBO R	12.3 ± 2.0	8.9 ± 2.5	
PBO NR	13.6 ± 4.0	12.5 ± 4.4	$F = 13.1$, df $= 1$, $p = .001$
Treatment			
Drug	13.3 ± 3.7	10.6 ± 4.8	
Placebo	14.7 ± 4.0	11.7 ± 4.3	$F = .3$, df $= 1$, $p = .6$
Covariate			
HDRS score at randomization			$F = 22.4$, df $= 1$, $p < .001$
Interaction			
Response by treatment			$F = .1$, df $= 1$, $p = .8$

Table 3. Estimate of Effect Size Contrasting CGI Change for Ultimate Responders and Nonresponders in Drug and Placebo and Between All Patients on Drug versus All Patients on Placebo

	Drug			Placebo			
	Ultimate Responder	Nonresponder	Effect size	Ultimate Responder	Nonresponder	Effect size	Drug-Placebo Effect Size
Responders week 1 CGI change							
Much and very much improved	13% (32/252)	6% (9/140)	0.24	21% (12/58)	7% (10/143)	0.31	0.01
Responders week 2 CGI change							
Much and very much improved	29% (74/252)	10% (14/140)	0.49	33% (19/58)	8% (12/143)	0.65	0.18

responder is significant ($F = 13.06$, $df = 1$, $p = .000$). However, there is no treatment effect ($F = .283$, $df = 1$, $p = $ NS) and no significant interaction.

An analysis of covariance (ANCOVA) contrasting the HDRS scores at 2 weeks of ultimate responders on drug and placebo was done. There is no difference between drug and placebo, suggesting that the change observed at 2 weeks is attributable to placebo effects ($F = 0.02$, $df = 1$, $p = $ NS).

Question 3

What is the effect size if ultimate responders and nonresponders are contrasted on the proportion rated responders at week 1 and week 2 in drug and placebo trials? What is the effect size if the proportion of all patients (ultimate responders and nonresponders combined) rated responders at week 1 and week 2 on drug is contrasted with all patients on placebo? The proportion of improved patients at week 1 and week 2 for ultimate responders versus nonresponders was used to estimate the effect size in the drug and placebo groups (Table 3). At week 1 on drug, 13% of ultimate 6-week responders are rated responders, and 6% as ultimate nonresponders. Using tables from Cohen (1988), the effect size is calculated as .24. A similar analysis in the placebo group contrasting responders at week 1 for the ultimate responder versus nonresponder resulted in an effect size of .31. When the difference in the proportion between all drug and placebo responders at week 1 was used, the effect size was .0. An identical analysis was done for week 2, with data yielding an effect size for drug of .49 and placebo of .65. The difference in the proportion between drug and placebo responders at week 2 resulted in an effect size of .18.

Using the responder versus nonresponder model contrasting outcome for ultimate responders versus nonresponders, the estimate of effect size in this sample is less than that of Katz et al. At week 1 the effect size is .24 and .31 for drug and placebo, respectively, and would fall in the small effect with no differences between drug and placebo. At week 2 within drug and

placebo the effect size is .49 within drug and .65 within placebo (contrasting ultimate responders versus nonresponders). This corresponds to approximately a 20% difference and is probably clinically detectable (Cohen 1988). However, when all drug data are contrasted to placebo data, the effect size of .18 is small and reflects a difference of 5% to 10% in the proportion of responders and is probably not clinically detectable.

DISCUSSION

Data included in the present report support the observation that early improvement is associated with a favorable prognosis. This has clinical relevance. However, these data do not support the hypothesis of Katz et al. (1987) that the effect of antidepressants can be observed within 1 week of initiating treatment. The patients studied by Katz et al. differed from our sample in that they were more severely ill inpatients. One way to reconcile the differences between the two analyses is to suggest that more severely ill inpatients would have a more rapid onset of beneficial effects. This seems counterintuitive. If inpatients had a lower placebo response rate, then the change observed by Katz et al. might be attributable to drugs. However, a review by Klerman and Cole (1965) suggests that placebo response rates of inpatient depressives varies widely, with nonchronic groups having a high placebo response rate and chronic groups a low placebo response rate.

The data presented strongly suggest that conclusions about the speed of onset of specific antidepressant effects cannot be determined without a placebo control. These data do not support the hypothesis that antidepressant effect can be observed within the first 2 weeks of initiating treatment. Periodically the ethical implication of including a placebo group is questioned if effective treatment exists (Rothman and Michels 1994). Support for this position comes from the viewpoint that a placebo is unethical because effective treatment is withheld. We suggest that a utilitarian assessment would conclude that a placebo control reduces cumulative morbidity because it is required accurately to assess the

efficacy of new treatment protocols as well as their time of onset.

REFERENCES

Bielski RJ, Friedel RO (1976): Prediction of tricyclic antidepressant response. Arch Gen Psychiatry 33:1479–1489

Casper RC, Katz MM, Bowden CL, Davis JM, Koslow SH, Hanin I (1994): The pattern of physical symptom changes in major depressive disorder following treatment with amitriptyline or imipramine. J Affect Disord 31:151–164

Cohen J (1988): Statistical Power Analysis for the Behavioral Science. New York, Academic Press

Coryell W, Coppen A, Zeigler VE, Biggs JT (1982): Early improvement as a predictor of response to amitriptyline and nortriptyline: A comparison of 2 patient samples. Psychol Med 12:135–139

Janicak PG, Davis JM, Preskorn SH, Ayd FJ Jr (1993): Principles and Practice of Psychopharmacotherapy. Baltimore, Williams & Wilkins

Joyce PR, Paykel ED (1989): Predictors of drug response in depression. Arch Gen Psychiatry 46:88–89

Katz MM, Koslow SH, Maas JW, Frazer A, Bowden CL, Casper R, Croughan J, Kocsis J, Redmond E Jr (1987): The timing, specificity and clinical prediction of tricyclic drug effects in depression. Psychol Med 17:297–309

Katz MM, Maas J, Frazer A, Koslow SH (1994): Onset of antidepressant activity, the structure of depression and multiple drug actions. Paper presented at ACNP, Puerto Rico

Klein DF, Gittelman-Klein R, Quitkin FM, Rifkin A (1980): Diagnosis and Drug Treatment of Psychiatric Disorders. Baltimore, Williams & Wilkins

Klerman GL, Cole JO (1965): Clinical pharmacology of imipramine and related antidepressant compounds. Pharmacol Rev 17:101–141

Liebowitz MR, Quitkin FM, Stewart JW, McGrath PJ, Harrison WM, Markowitz JS, Rabkin JG, Tricamo E, Goetz DM, Klein DF (1988): Antidepressant specificity in atypical depression. Arch Gen Psychiatry 45:129–137

McGrath PJ, Quitkin FM, Harrison W, Stewart JW (1984): Treatment of melancholia with tranylcypromine. Am J Psychiatry 141:286–2894

McGrath PJ, Rabkin JG, Stewart JW, Harrison W, Quitkin FM, Markowitz J (1985): Placebo-controlled study of mianserin in depressed outpatients. Neuropsychobiology 14(3):128–132

McGrath PJ, Stewart JW, Nunes EV, Ocepek-Welikson K, Rabkin JG, Quitkin KM, Klein DF (1993): A double-blind cross over trial of imipramine and phenelzine for outpatients with treatment refractory depression. Am J Psychiatry 150(1):118–123

Quitkin FM, Liebowitz MR, Stewart JW, McGrath PJ, Harrison W, Rabkin JG, Markowitz J, Davies SO (1984): L-Deprenyl in atypical depressives. Arch Gen Psychiatry 41:777–781

Quitkin FM, Rabkin JG, Markowitz JM, Stewart JW, McGrath PJ and Harrison W (1987): Use of pattern analysis to identify true drug response: A replication. Arch Gen Psychiatry 44:259–264

Quitkin FM, McGrath PJ, Stewart JW, Harrison W, Wager SG, Nunes E, Rabkin JG, Tricamo E, Markowitz J, Klein DF (1989): Phenelzine and imipramine in mood reactive depressives: Further delineation of the syndrome of atypical depression. Arch Gen Psychiatry 46:787–793

Rothman KJ, Michels KB (1994): The continuing unethical use of placebo controls. N Engl J Med 331(6):394–398

Stassen HH, Angst J, Delini-Stula A (1994): Severity at baseline and onset of improvement in depression. Meta-analysis of imipramine and moclobemide versus placebo. Eur Psychiatry 9:129–136

Stewart J, Quitkin FM, Fyer A, Rifkin A, McGrath P, Liebowitz M, Rosnick L, Klein DF (1980): Efficacy of desipramine in endogenmorphically depressed patients. J Affect Disord 2:165–176

Stewart JW, Quitkin FM, McGrath PJ, Liebowitz MR, Klein DF (1981): Efficacy of desipramine in mildly depressed patients: A double-blind placebo-controlled trial. Psychopharmacol Bull 17(1):159–161

Winokur G (1980): What to do? Or what do we owe our residents? Biological Psychiatry 15:599–611