
PERSPECTIVES

Enhancing the Technology of Clinical Trials and the Trials Model to Evaluate Newly Developed, Targeted Antidepressants

Martin M. Katz, Ph.D., Uriel M. Halbreich, M.D., Charles L. Bowden, M.D., Alan Frazer, Ph.D., Roger M. Pinder, Ph.D., A. John Rush, M.D., David P. Wheatley, M.D., and Barry D. Lebowitz, Ph.D.

Concern about disappointing results from recent multi-center trials of new antidepressants prompted several ACNP workshops on "improving the technology of clinical trials." The workshops focused on technical problems, such as patient screening, reliability of clinical ratings, and the role of the placebo control. They aimed to determine how to more effectively apply the current clinical trials model for evaluating antidepressant drugs. The problems confronting the field of clinical trials, however, extend beyond technology. They also included conceptual issues concerning changes in the understanding of depressive disorders and of the multiple actions of antidepressant drugs. Such problems have been further complicated by the rapidly changing field of drug development itself, which is continually refining the targeting of new antidepressant agents. Drugs are increasingly being developed to try to change specific behavioral facets more rapidly and may be less likely, therefore, to act initially on "whole" disorders. To address such issues, a symposium was held in Rhodes in

2000 that focused on such conceptual changes with the goal of developing recommendations to revise the clinical evaluation model. Its purpose was to integrate new knowledge on depression and the mechanisms of action of antidepressant drugs toward developing more efficient methods of drug development. Since the evaluation process will eventually require changes in governmental policy, senior staff from the National Institute of Mental Health (NIMH) and Food and Drug Administration (FDA) participated as well as members of academia, industry and clinical practice. Recommendations for altering clinical trial methodology were made in four areas: patient selection, methodology of evaluation, measuring onset of action, and FDA and NIMH perspectives on current practice. This article discusses these four areas and presents the consensus of the panel participants.

[*Neuropsychopharmacology* 27:319–328, 2002]

© 2002 American College of Neuropsychopharmacology.
Published by Elsevier Science Inc.

KEY WORDS: Antidepressant; Clinical trial; Antidepressant drug development; Clinical methodology; Componential behavioral approach; Onset of drug action

From the Department of Psychiatry (MMK, CLB), and Department of Pharmacology (AF), University of Texas Health Science Center at San Antonio, San Antonio, TX, Department of Psychiatry, State University of New York at Buffalo (UMH), CNS & Cardiovascular Division, Organon, Inc. (RMP), Department of Psychiatry, University of Texas Southwestern Medical Center at Dallas (AJR), FRCPsych, Psychopharmacology Research Group (DPW), and Adult & Geriatric Treatment & Preventive Interventions Research, National Institute of Mental Health (BDL).

New methods of drug development based on recent research on mechanisms of action have resulted in a series of new putative antidepressant drugs. Many of the new drugs are aimed at producing greater efficacy, re-

Address correspondence to: Martin M. Katz, Ph.D., Department of Psychiatry, University of Texas Health Science Center, 6305 Walhonding Road, Bethesda, MD 20816. Tel.: (301) 229-1175; Fax: (301) 229-3784; E-mail: mkbsk@Worldnet.ATT.net

Received July 24, 2001; revised January 20, 2002; accepted March 8, 2002.

Online publication: 3/12/02 at www.acnp.org/citations/Npp031202266.

ducing adverse events, and lessening the complexity of dosing regimens. In addition, attention is currently focused on accelerating the onset of drug action to generate more rapidly acting antidepressants. Unfortunately, recently conducted multi-center clinical trials of some of these agents encountered serious technical problems and did not produce the results anticipated from pre-clinical studies. Although this may be partly due to limitations in extrapolating pre-clinical data to patients, it may also be that there are significant constraints in the current clinical trials model that make it difficult to demonstrate efficacy.

The current clinical trials model is now 30 years old. The essential principles of randomization, double blind evaluation, and a placebo control remain sound. But other important aspects of the research design and the methodology are no longer sufficient or well suited to the testing of these new drugs. In effect, such deficiencies in current trial methodologies make it increasingly difficult to demonstrate antidepressant efficacy of these new drugs.

One important reason the current model has proved inadequate is that it was initially applied to hospitalized depressed patients. Today new drugs are tested almost exclusively on outpatient samples, many of whom are symptomatic volunteers, responding to advertisements. Depressed outpatients have milder depressive symptomatology than the predominantly inpatient samples used to test the first established drugs. Patients with milder severity provide a narrower range of possible change, and produce a greater number of placebo responders (Fisher et al. 1965; Bowden et al. 2000; Khan et al. 2002).

As a consequence of the smaller difference between active treatment and placebo treatment, larger patient samples and/or more sensitive clinical methods are needed to detect differences between drug and placebo. The most commonly utilized scales, the Hamilton Depression Rating Scale (HDRS) (Hamilton 1960) and the Clinical Global Impression Scale (CGI) (Guy 1976), were designed to measure changes in overall severity of the disorder. These scales are not sufficiently sensitive to assess the specific behavioral changes brought about by the newer drugs (Montgomery and Asberg 1979). Further, intensive evaluation of symptomatology and behavioral change are carried out infrequently with outpatients, making evaluation of rapidity of response problematic. Consequently, it appears important to improve the technology applied to assessing the specific actions, onset, and efficacy of the newer drugs.

Acknowledging the increased complexity of the goals of clinical trials and the need to improve trial procedures, the designers of the Rhodes panel identified four critical problem areas to resolve: (1) the nature of the patient sample; (2) behavioral methods and analyses used for assessing treatment-induced improvement

and recovery; (3) the lack of standards for determining onset and speed of action; and (4) the failure to integrate advances in knowledge of depression and the antidepressants into current clinical trials.

PATIENT SELECTION

New drugs for depression are being designed to be effective across the entire range of severity of the disorder, to minimize adverse events, and to effect a rapid onset of therapeutic action. In the United States, the targeted population for studies is now those patients currently being treated in an outpatient psychiatric setting. The established system for diagnosing the disorder is the DSM-IV. The HDRS is the established instrument for determining whether the patient is sufficiently depressed to enter a clinical trial. Major reasons for failure of some recent trials are the inclusion of patients as subjects who are "symptomatic volunteers", patients who barely meet the minimal level of qualification, or those who are admitted because of a tendency on the part of clinicians, consciously or unconsciously, to lower their threshold for perceiving symptoms. Patients or symptomatic volunteers in whom the condition of depression is mild to "marginal" tend to have a high placebo response rate (Khan et al. 2002) in part because they are more responsive than more severely ill patients to non-specific psychosocial supports associated with clinical studies. Therefore, their presence could seriously undermine the ability to show efficacy. Indeed, in most inconclusive trials, the main reason for lack of separation between placebo and active drug has been purported to be high placebo rate (Uhlenhuth et al. 1997). Patient selection also excludes comorbid conditions that are common with depression, especially when placebo-controlled trials are being conducted. Thus, general medical conditions, Axis II, other Axis I disorders, e.g. panic disorders, even GAD, bulimia, anorexia, substance abuse, are all excluded. Secondary problems noted in this area were the selection of patients who, for various reasons, were likely to drop out early, the refusal of otherwise qualified patients to participate in the trial, and the deficiencies of particular study sites in discriminating active drug from placebo treatments. Also, current protocols include subjects who have failed to respond to multiple prior treatments in the current episode and thus may be treatment-resistant.

Patient selection has been seen as the principal source of error in recent trials, so much attention has been directed toward solving this critical problem.

The participants in the panel recommended the following to increase the rigor and quality of the screening process:

1. Eliminate non-declared patients or "symptomatic volunteers" from clinical trials.
2. Refine the rating criteria. The items of the screening instrument, the Hamilton scale, are heavily weighted with anxiety and somatic symptoms. These symptoms identify patients who are pathological and, therefore achieve a qualifying score, but who may not be primarily depressed. The Hamilton screening score should therefore be supplemented, at the minimum, by requiring the patient receive at least a "2" on Item No. 1 of the scale, "depressed mood" to qualify.
3. Separate the screening and evaluative procedures from the administration of treatment. This would separate the treater from the screening evaluator. It may require introducing another professional into the study or including a video procedure that will permit an independent evaluator to rate the patient's progress throughout the course of treatment.
4. Eliminate patients whose initial high score on severity can be traced solely to an immediate stressful event. It may be necessary to require that a prospective patient have the depression for at least several weeks.
5. In order to reduce the proportion of early dropouts, eliminate patients whose occupation will be seriously impacted by one of the expected side effects of the experimental drug, e.g., the effect of "drowsiness" on operators of construction or farm equipment, "dry mouth" on salespeople, etc.
6. Reduce the number of patients who qualify but who decline because of the presence of a placebo control. More severely disturbed patients may avoid controlled studies that include a placebo control. Randomized placebo treatment should therefore be relatively short in length. Patients should be promised treatment with an already established drug if they do not respond to the study treatment within three to four weeks. Other suggestions include utilizing or offering limited forms of counseling as an alternate intervention to all participants, as is done frequently in substance abuse treatment trials.
7. Eliminate study sites where it can be determined that the large majority of patients at that site cluster around the minimal Hamilton score required for admission to the study. Recent studies (Niklson et al. 1997) have reported that certain study sites were unable to show differences between the "active" treatment control (imipramine) and placebo. The authors did not find differences between the sensitive and insensitive sites in regard to adequacy of raters, a generally poor treatment environment, or the inclusion of too many patients who were not adequately depressed. They did find evidence that those sites which identified superiority of imipramine over placebo generally established and maintained better quality therapist-patient relationships, although this point remains controversial, e.g., both placebo and active treatment response were equally enhanced in the better quality relationships. However, it is also possible that patients at the sites were inadequately screened. An additional complication is the possibility that a study site is inconsistent across different trials. Informal reports from two major companies on this issue are contradictory. Consequently, it was proposed that study sites in a major trial of a new antidepressant be required to demonstrate their capacity to distinguish an active control from the placebo control, either in their investigational history or in the preliminary stage of the study proper. Consideration should be given to inclusion in the final efficacy analysis only of those sites that distinguish active drug control from placebo by two points or more on the HDRS at the end of the trial.
8. The effect of the "run-in" period on dropouts: Recent analyses indicated that the inclusion of a "run-in" period in the trial, e.g., one week on placebo prior to randomization, increased the dropout rate, but did not affect the size of the difference between the drug and the placebo control during the randomized phase (Trivedi and Rush 1994). Nevertheless, the participants deemed it useful for several reasons to retain the "washout" period, with or without a pill placebo. First, it provides an initial test of the stability of the patient's depressed state. Second, it will eliminate some early placebo responders. Third, it will provide time to wash out most of the effect of prior medications, e.g., improvement in depressive symptoms could result in this situation from discontinuation of medication rather than from effects of the new treatment.

In summary, a number of factors associated with the characteristics of the patients and the screening criteria appear to interfere with generating proper evaluation of a new treatment. In light of such issues inappropriate selection appears to be a major source of error in current clinical trials of antidepressants. It appears that pressure to accumulate large enough samples for study prompts the acceptance of many subjects who would ordinarily not qualify for treatment with the new agent. Other factors cause some patients to drop out despite their being fully qualified. A major step forward toward obtaining clear results in the efficacy of a drug will be to direct significantly greater attention toward selecting the appropriate patients for study and to choose study sites well-qualified to conduct clinical trials of new agents. In the next section are suggestions on how to remove certain obstacles blocking that accomplishment, and how to assemble a more representative sample of patients.

METHODOLOGY OF EVALUATION

One of the principal causes for the failure of recent multicenter outpatient trials is the limited methodologic ap-

proach to measuring behavioral change that is applied in the conventional clinical trial. An aim of new drug development, in addition to improving (global) efficacy for a disorder, is to focus as well on specific symptoms, e.g., reducing anxiety or sleep disturbance or anhedonia. The goal also is to initiate more rapid improvement. So, behavioral methods must go beyond the assessment of overall severity and measure the speed of the effects of drugs on the major components of the affective disorders. Such components include depressed mood, anxiety, hostility, psychomotor dysfunction, cognitive impairment and somatic symptoms. When differences between the experimental and the control treatment are large, use of the HDRS and Montgomery-Asberg Depression Rating Scale (MADRS) (1979) items to measure these components provide crude estimates, which can at times be effective discriminators. In patients with milder forms of the disorder, overall change can be expected to be smaller while larger effects on critical aspects of the disorder are more difficult to demonstrate. Therefore, more sensitive measures may be required to detect drug-placebo differences. Such measures will be discussed in the next section followed by specific recommendations of methods that can increase the sensitivity to detecting change and overall outcome.

Conceptions of Depression and the Mechanisms of Drug Action

Changes in the concept of depression and how the antidepressant drugs bring about therapeutic effects influence our definitions of improvement and outcome. For example, depression is now classified into categories, unipolar and bipolar, and viewed as multifaceted, comprised of major affective, cognitive, psychomotor and somatic components (Maas et al. 1991). Antidepressants are no longer seen as "specific" to depressive disorders. They have also been shown to be effective for generalized anxiety disorder (even early on, imipramine was demonstrated to be more effective than the then leading "anti-anxiety" drug (Kahn et al. 1979)), selected phobias, and obsessive-compulsive disorders (Insel 1991). It appears that reductions in anxiety and hostility are the initial signs of response in depressed patients to the tricyclic drugs, preceding the reduction of depressed mood (Haskell et al. 1975; Katz et al. 1987, 1991). Such evidence supports the results of earlier research by Kielholz and Poldinger (1968) and Carlsson et al. (1969), indicating that the tricyclic drugs have multiple actions, i.e. both tranquilizing and mood-elevating effects. These findings encourage clinical investigators to apply componential measures appropriate to the recording of such effects when assessing improvement and response in the testing of new drugs.

To increase the sensitivity of current evaluative measures, the panel recommended the following:

1. Include self-report instruments that record the patient's distress directly. When the disorder is severe, the patient's symptoms are more pronounced, and thus, more observable. The self-report may then add little to the interview-based clinical rating. When the symptoms are mild, however, their comparative severities are more difficult for the rater to estimate from the brief interview period of observation. Some affective states, such as hostility, are more likely to be reported and more sensitively recognized by the patient than by trained raters (Weissman et al. 1971; Katz et al. 1987). The direct self-report test can therefore add significantly to the clinical rating procedure. A number of such tests have frequently been used in trials as secondary outcome measures. Ones with demonstrated validity in outpatient trials are: the Symptom Checklist-90 (SCL-90) (Derogatis et al. 1974), the NIMH Mood Scale (Raskin et al. 1967), and the Inventory of Depressive Symptomatology-Self-Report (IDS-SR) (Rush et al. 1986, 1996).
2. Adopt a "componential" approach to measurement: self-report inventories are usually used to simply reinforce findings from the more established clinical rating indices, but they can also be used to extend the evaluation to cover changes in the major components of the disorders. The most prominent example is in the assessment of changes in the components of anxiety and hostility, key aspects of the depressive disorder. The hostility variable, viewed as a core element in several theories of the genesis and nature of the depressive disorders (Freud 1959; Abraham 1949), is not even included as a symptom in the HDRS and in the MADRS. This is so despite the evidence that hostility has been shown to be a very sensitive index of tricyclic drug action in treatment-responsive patients (Fava et al. 1986; Katz et al. 1987).
3. Combine clinical methods to provide the primary outcome measures: Problems arise as to how methods supplementary to the primary outcome measure can be combined or integrated into the efficacy equation. It is easy to envision trials in which the central issue remains the severity of the total disorder, but in which the trial can also evaluate whether a new drug is effective in reducing the severity of major psychopathologic components. Demonstration of significant efficacy in reducing components of the disorders in the early investigation of the actions of the tricyclic drugs and the SSRI drugs would have indicated their significant impact on anxiety and hostility. That could have led more quickly to the drugs being applied to disorders other than depression, resulting in great therapeutic and economic benefits. Discovering how these methods can be combined to meet the objective of devising primary outcome indices and measures of the major components requires psychometric analysis of large sample data in which

the methods have been applied alongside the conventional severity rating scales. Examples of systems based on combining several diverse test procedures are found in Raskin et al. (1967), and in Katz et al. (1984). The values of the componential approach in meeting current problems of evaluation are further examined in the section on measuring onset and speed of action of the new drugs.

4. Independent evaluation using video methods: The use of a standardized video procedure can separate the clinical evaluator from the treater. It permits objective evaluation of treatment progress. A system currently in use, the Video Interview Behavior Evaluation Method (VIBES) (Katz et al. 1989) includes a standard interview and a set of specially adapted rating instruments in addition to use of the established HDRS. Specially devised rating scales also take advantage of the strengths of video observation by including motor and cognitive tasks within the interview and focusing on expressive and social behavior. Although its use for drug research has been demonstrated, the technique still requires extra time and expense, more than is currently devoted to evaluation in clinical studies. However, as briefer versions of the video scales become available, they should provide further, perhaps more important additional information more conveniently.
5. Functional measures: In accord with the renewed interest in "remission", i.e., the virtual absence of symptoms and restoration of social function, as an outcome index, as against response (>50% reduction in baseline symptomatology), recent evidence shows a place for measures of social functioning and quality of life (Dubini et al. 1997; Healy 1998; Bosc et al. 1997). Studies have shown such measures to be sensitive to drug actions when symptom severity in itself was not (Dunbar et al. 1991). Signs of social impairment in outpatient depressives can be subtle. However, the positive results derived from such studies are probably due to the fact that even with mild symptoms, serious social impairment can occur in outpatients who have had a chronic, rather than simply episodic, course (Miller et al. 1998). The measures utilized in this area include the Social Adaptation Self-Evaluation Scale (SASS) (Bosc et al. 1997) and the Social Adjustment Scale (SAS) (Weissman and Bothwell 1976) and other social functioning scales. This area is being pursued more intensively in current clinical trials.
6. Further efforts to increase the reliability and validity of HDRS and other global severity scales: New techniques for training developed with the explicit aim of reducing variance among raters within and across institutions. Advances in digitalization and teleconferencing have made the new video approaches applicable to the simultaneous training of

raters across multiple sites (Stahl and Shayegan 2001). These procedures further strengthen the HDRS or other clinical symptom scales as overall measures, significantly reduce rater variance, and should contribute to reducing the variability of results seen across multi-site studies.

In summary, detecting differences between the actions of the new antidepressants and placebo during the course of treatment in samples that are limited to mild and moderately severe depression requires a significant increase in the sensitivity of the methods of evaluation: (1) The HDRS, which has been quite successful as a discriminator in detecting antidepressant actions when the population is in the moderate to upper range of severity, is not adequate per se for these additional purposes. (2) Additional methods, which focus on the symptom or behavioral characteristics of the outpatient depressive, need to be applied. These include already established self-report instruments, e.g., the SCL-90, and, where possible, new methods designed to enhance sensitivity in detecting early specific drug actions. Greater emphasis in future trials should be placed on measuring drug actions on the major components of depression, as well as on the severity of the whole disorder. This componential approach, which identifies certain affective, cognitive and psychomotor elements as primary facets of the disorder may require the systematic combining of clinical rating, self-report and objective performance measures. (3) Measures of social behavior and adjustment to assess "remission of functioning" should be included. Since adaptation or development of new methodology for evaluation is not an area currently being supported by the pharmaceutical companies, the NIMH can play a very important role. The Institute can stimulate, assist and facilitate the funding of applied research. Development and testing of new methods of evaluation has been an area of much neglect.

ONSET AND SPEED OF ACTION OF ANTIDEPRESSANTS

There are now many different types of drugs that ameliorate depression. In general, such drugs are currently hypothesized to act by enhancing neurotransmission mediated by norepinephrine (NE) and/or serotonin (5-HT). The ways they do this may vary, but irrespective of their specific mechanisms, most have acute, fast-developing effects on biogenic amine systems.

Even though the drugs have immediate effects on norenergic or serotonergic neurons, there is agreement that their maximal or optimal therapeutic effects in depression may be delayed for six weeks or longer. Somewhat more controversial is whether their onset of efficacy, defined as the first time when behavioral improvement becomes clin-

ically significant or “visible”, is also delayed. A commonly held belief is that onset is delayed for two to three weeks after treatment begins. Contrarily, Kuhn (1958), the discoverer of imipramine, reported strong clinical effects within the first week, approaching remission in some patients. Later, placebo controlled studies were equivocal and unable to clearly support these early findings. Unfortunately, there are few, if any, prospective, appropriately designed clinical studies that have addressed this issue (Katz 1998).

Until recently, no study has been specifically designed to address the timing issues of antidepressant drug action. Most of our current understanding of the course of recovery for depression stems from drug trials involving only weekly assessments of efficacy measures. Recently, the application of new statistical procedures to the problem of estimating onset has led to unusual results. Conventional repeated measurement analyses of variance, or mixed linear models in the case of unbalanced data, generally demonstrate a statistically significant advantage of active compounds over placebo after three weeks of treatment. In contrast, survival-analytical methods suggest that among responders, onset of action occurs in more than 70% of cases within the first two weeks of treatment, with an early onset of action being highly predictive of later response. (Angst and Stassen, personal communication; Stassen et al. 1993, 1997; Coryell et al. 1982; Boyer and Feighner 1994).

When a new drug designed to increase effectiveness and to act more rapidly is produced, the attempt to evaluate its efficacy and onset in one clinical trial is obviously complicated. Some new drugs have yielded evidence that they act more rapidly than the established ones. But the evidence has not been convincing. Demonstrating a statistical difference between drug and placebo at one week of treatment is, for statistical and other reasons, not adequate in itself to establish one drug as having a faster action.

Stassen et al. (1993) developed a more persuasive approach to determining onset and speed of action by analyzing change within each of the treated patients. They defined a 20% reduction on the HDRS as a sufficiently stable index of improvement, with the additional requirement that the improvement when measured on subsequent days throughout the course of treatment not fall below 15%. Use of this procedure permits the derivation of an ordinal estimate of the onset of drug action. For example, the use of the median time point, i.e., the time point at which 50% or more of patients achieve 20% improvement that is then sustained through the course of treatment, can be calculated for each treatment studied. This is a promising approach for determining “onset.”

To satisfactorily determine time to onset, several issues require resolution. “Improvement”, for example,

must be distinguished from “full response” and “remission.” Stassen et al. (1993) adopted the 20% criterion for the HDRS as the index of improvement, and the time point at which 50% improvement is achieved, as the definition of response. Montgomery (1995), for other reasons, proposed 25% as the initial threshold for improvement. The design of a trial to estimate onset requires scheduling earlier and more frequent points of evaluation than the conventional week by week approach. Recent trials are adopting a twice-weekly approach for a minimum of the first three weeks during treatment, to be followed by weekly assessments. Adequate dosage of the drug, representative of how the drug will be administered in practice, must be achieved for all patients early in treatment.

Much has been written recently about the substantive and statistical requirements for applying sound methodology to this problem (Gelenberg and Chesen 2000; Leon 2001). In a recent review of the issues of onset measurement, Katz et al. (1997), included significant changes in the major components of the disorder, e.g., anxiety, depressed mood, psychomotor function, as evidence of therapeutic action. Thus, significant reduction in these components also qualified as indices of onset.

Some issues remain controversial. Angst and Stassen (personal communication) note that to insure proper assessment of the new drug, sufficient dosage must be initiated early in the treatment. Many studies (because of extended initial periods for graduating dosage) have not provided a fair test of possible early onset of the treatment. The inadequate dosages weakened any likelihood of establishing onset for any specific drug treatment. The “wash out” period is an additional complication since not taking it into account can confound estimates of the time of onset of the new drug. Laska and Siegel (1995) have proposed techniques to determine the onset of action where only a proportion of patients respond to a treatment. Statistical techniques, including survival curves, are provided for distinguishing the onset of response in treatment-responders, from that estimated in the study sample as a whole. Again, such approaches need to be incorporated into clinical trials.

Time of onset and the speed of action of antidepressants are issues that are now front and center for the pharmaceutical companies. For the reasons reviewed here, the conventional model for clinical trials will not be sufficient to characterize the effects of such drugs. Investigators must do more than simply adopt more frequent assessment of drug actions. Because the FDA is not yet in a position to provide standard criteria for establishing that one drug acts more rapidly than another, there is a need for clinical investigators, possibly with the support of NIMH, to conduct the necessary studies that will clarify the needed methodology and criteria.

In summary, the statistical and methodologic requirements for a study aimed at determining definitive

estimates of onset and speed of action of an antidepressant are not yet resolved. However, the minimal study requirements as viewed at this time are extended from those presented earlier by Prien et al. (1985). They are:

1. Criteria-based operational definitions of improvement, response and remission.
2. Measurement of the severity of the disorder and its major components, including anxiety, hostility, psychomotor dysfunction and cognitive impairment.
3. Statistical criteria for determining thresholds of improvement and full response.
4. Including in the design appropriately frequent assessment points during the early phase of treatment.
5. Ensuring proper controls; this would include, depending on the specificity of the question, placebo and /or active treatment control.
6. Ensuring that adequate dosage, representative of how the drug will be administered in practice, is achieved for all patients.

FDA AND NIMH POLICY PERSPECTIVES ON CLINICAL TRIALS OF ANTIDEPRESSANT DRUGS

FDA Perspectives

Under the authority of the Food, Drug, and Cosmetic Act, regulations are written to provide operating guidelines and working definitions for terms such as "effective." More than one well-designed and adequately controlled study that is significantly positive by itself is the standard for establishing efficacy. Generally, Congress implied that "more than one" meant at least two studies. There are several possible adequate controls for clinical trials of drugs in the treatment for depression. The Division of Neuropharmacologic Drug Products of the FDA requires that a drug must show superiority over a control. Clinical trial designs that suggest non-inferiority to an active control are not acceptable as sufficient evidence for approval. Most trials of drugs for the treatment of depression include both a placebo and an active control. It is very common that both the active control and study drug show equivalence, while neither shows superiority over placebo. This can be for a myriad of well-recognized reasons; however, the controversy over the use of placebo in trials of drugs for the treatment of depression usually revolves around issues of placebo treatment in the face of the risk of suicide.

There appears to be no difference in the rate of either completed or attempted suicide in outpatients treated with placebo vs. either active comparators or study drugs in outpatients involved in clinical trial research. In the study of 19,639 patients in clinical studies of antidepressant drugs, annual rates of suicide and attempted suicide respectively, were 0.8% and 2.8% for

investigational drugs, 0.7% and 3.4% for active comparators, and 0.4% and 2.7% with placebo (Khan et al. 2000). However, actively suicidal patients are not included in clinical trials. What these data show are that the rates at which depressed patients who are not suicidal at study entry who go on to attempt or complete suicide in short term studies is not affected by drugs that have been tested or approved up to this point in time. Enrollment in a placebo arm of a clinical trial for the treatment of depression appears to be equally as life saving (or as ineffective at preventing progression to suicidal behavior) as approved active or investigational treatment. Given this data, the Division of Neuropharmacologic Products will most likely continue to require that drugs used in the treatment of depression show superiority over something that is at least as good as placebo. Patients enrolled in the placebo groups in these studies also showed a 30.9 per cent reduction in depressive symptoms vs. 40.7% for investigational drugs and 41.7% for active comparators. These and related studies provide conclusive evidence that in the case of clinical trials of depression, placebo treatment does not constitute "no treatment."

As the medical community learns more about the disease states that lead to depressive episodes, one sees the need for more discriminating choices of treatment options. Until recently, drugs used for the treatment of depression have been labeled as "antidepressant" drugs. The patients enrolled in the clinical trials that support claims of efficacy for these drugs have had diagnoses most closely resembling DSM-IV Unipolar Major Depressive Disorder (either recurrent or single episode). It is now clear that depression has various forms that may respond differently to treatment based on the overall diagnostic perspective of the depressive episode. For example, drugs that are effective and safe for the treatment of depressive episodes in patients with unipolar depression may be neither safe nor effective for the treatment of patients with depressive episodes associated with bipolar disorder. The Division will now be labeling drugs by the indication in which they were studied. This change is consonant with strong evidence that some drugs are beneficial in more than one psychiatric disorder, and that depression is not a unidimensional condition, e.g., the substantial differences existing between recurrent unipolar and bipolar depressions (Bowden 2001).

The treatment community is often interested in potential differences in various drugs used to treat depression. These interests are driven in part by desires to prescribe drugs with the fewest associated adverse events, faster onset of action, or lowest cost relative to their efficacy. Pharmaceutical companies have reasonable interests in conveying information on differences in the spectrum of actions of one drug compared with others in a group. Competitive claims may now be allowed if

they are supported by well-designed adequately controlled data. Such claims continue to be closely scrutinized by the agency.

The difficult part about regulatory work is that it combines both the law and science. In the United States, law is based on precedent, and precedent is best applied in an environment where things do not change. By contrast, in science, our views and understanding are constantly changing. Regulatory decisions acknowledge that the way that we look at diseases and the human condition actually does change. Therefore, regulatory guidelines often change over time. Regulatory changes can be both welcome and unsettling to the competitive world of drug development. When the FDA does make changes in how drugs are reviewed, especially antidepressants, they have been extensively considered. The Agency seeks input from industry, patients, and the scientific community to ensure that its policies regarding depression (as well as other indications) are at once, flexible, consistent, and scientifically sound because changes in policy are unsettling in the competitive world of drug development. The Agency looks at such changes long and hard, knowing that once policies change, new regulations will be with the field for a long time. Many recent European studies have not required placebo controls. This explains why some drugs which have been tested and rated to be effective by some European regulatory bodies have not been approved in the United States. The current policies of the Agency for the reasons previously outlined will remain in force into the foreseeable future. The FDA Neuropharmacological Drug Products Division policy change, to acknowledge differentiation between unipolar and bipolar depression, and differentiation in their responses to various antidepressant treatments, was based on adequate supporting evidence.

However, advances produced over the past three decades will require changes in policy that go beyond that of subtyping classification. At some point, the community of investigators and the FDA need to discuss whether a new set of criteria and a new model for conducting clinical trials designed to accommodate these advances is necessary. In view of the complexity of the major aims of clinical trials with the emphasis of developing more rapidly acting drugs, coupled with the rate of failure of multisite trials, we have to ask whether we may be warranted in making certain evolutionary changes even without unanimous accord of all interested parties.

NIMH Perspectives

In another context, the NIMH is now experimenting with trials that "simulate" actual practice, including the combination of treatments, as compared with conduct of randomized controlled clinical trials (Norquist et al.

1999). NIMH sees the outcome of the classic Phase III efficacy trial as an important intermediate step in the treatment development process, not the conclusion of that process. Phase III trials done for registration purposes may exaggerate the results of a given treatment's efficacy over that which can be expected when it is applied in actual practice. (An equally strong case can be made that the constraints that the clinical trial places on the treater results in the treatment being less effective than it would be in actual practice). The new public health orientation in clinical trials at NIMH was made possible by the expansion of resources and budget increases received in the last few years. The trials are an extension of current efforts in smaller scale studies, pilot studies and method development. Their position makes sense when trying to apply the results of clinical research to the broad goal of making the results immediately available to the public and when responding to the demands in the population for improvements in health care, generally.

However, the role of NIMH in the past has generally been to provide support toward the solution of technical and methodologic problems that serve as obstacles to progress in the development of new treatments. There is no more important an area than refining the procedures whereby new, innovative treatments are evaluated and made available to those who treat patients with affective disorders and the major psychoses. The NIMH could facilitate this process by encouraging the development of focused conferences on critical issues, and increasing the effort, through a more active funding program, to improve the methodology of evaluation.

Because methodology is not the most popular area for research and because the drug companies and peer review committees are slow to provide funds for research on methodology, NIMH should also make clear to investigators that support for such research can be obtained. Further, since the FDA staff is short-handed in psychopharmacology, the NIMH can serve as a vehicle for bringing a range of experts together with FDA officials and drug company investigators. Such meetings would be aimed at determining the sequence with which new needed changes can be integrated into the trials model, without disrupting long years of precedent and the balance of competition in the drug field.

Finally, NIMH officials and their advisors must be reminded or repeatedly be made aware that drug treatments are our primary defense against such disabling mental disorders as bipolar depression, mania and schizophrenia. A number of large, very expensive multi-center trials of new drugs have failed recently. These failures signal, in part, the fact that a once successful mechanism for evaluating the efficacy of new treatments is now fading in value, and not capable of meeting the more complex aims of evaluating newly

developed drug treatments. This is a major problem for the pharmaceutical industry. It is also, however, critical for the NIMH and the FDA to contribute their unique resources to resolving these problems, and subsequently, to facilitating the development of new, effective treatments for the vast number of mentally disordered patients.

ACKNOWLEDGMENTS

Based on Panel and Roundtable Conference, "Improving the Technology of Clinical Trials of Antidepressants," held at the 2nd International Congress on Hormones, Brain and Neuropsychopharmacology, Rhodes, Greece, July 2000.

The authors wish to acknowledge the contributions of Paul Andreason, M.D., at the Rhodes Conference and his assistance in writing the section on "FDA Policy Perspectives on Clinical Trials of Antidepressant Drugs."

REFERENCES

- Abraham K (1949): Selected Works of Karl Abraham. London, Hogarth Press
- Bosc M, Dubini A, Polin V (1997): Development and validation of a social functioning scale, the Social Adaptation Self-evaluation Scale. *Eur Neuropsychopharmacol* 7: S57–S70
- Bowden CL, Calbrese JR, McElroy SL, Gyulai L, Wassef A, Petty F, Pope HG Jr, Chou JC-Y, Keck PE Jr, Rhodes LJ, Swann AC, Hirschfeld RMA, Wozniak PJ (2000): A randomized, placebo-controlled 12-month trial of divalproex and lithium in treatment of outpatients with bipolar I disorder. *Arch Gen Psychiatry* 57:481–489
- Bowden CL (2001): Strategies to reduce misdiagnosis of bipolar depression. *Psychiatr Serv* 52:51–55
- Boyer WF, Feighner JP (1994): Clinical significance of early non-response in depressed patients. *Depression* 2:12–35
- Carlsson A, Corrodi H, Fuxe K, Hokfelt T (1969): Effects of some antidepressant drugs on the depletion of intraneuronal brain catecholamine stores caused by 4, alpha-dimethyl-meta-tyramine. *Eur J Pharmacol* 5:367–373
- Coryell W, Coppen A, Zeigler VE, Biggs J (1982): Early improvement as a predictor of response to amitriptyline and nortriptyline: a comparison of 2 patient samples. *Psychol Med* 12:135–139
- Derogatis L, Lipman R, Rickels K, Uhlenhuth E, Covi L (1974): The Hopkins Symptom Checklist (HSCL): A measure of primary symptom dimensions. In Pichot P (ed), *Psychological Measurements in Psychopharmacology: Modern Problems in Pharmacopsychiatry*, Vol 7. Basel, S Karger, pp 79–110
- Dubini A, Bosc M, Polin V (1997): Noradrenaline-selective versus serotonin-selective antidepressant therapy: differential effects on social functioning. *J Psychopharmacol* 11:S17–S23
- Dunbar GC, Cohn JB, Fabre LF, Feighner JP, Fiebre RR, Mendels J, Shrivastava RK (1991): A comparison of paroxetine, imipramine and placebo in depressed outpatients. *Br J Psychiatry* 159:394–398
- Fava GA, Kellner R, Lisansky J, Park S, Perini GI, Zielezny M (1986): Hostility and recovery from melancholia. *J Nerv Ment Dis* 174:414–417
- Fisher S, Lipman RS, Uhlenhuth EH, Rickels K, Park LC (1965): Drug effects and initial severity of symptomatology. *Psychopharmacologia* 7:57–60
- Freud S (1959): Mourning and melancholia. In *Collected Papers*, Vol. 4. New York, Basic Books
- Gelenberg AJ, Chesen CL (2000): How fast are antidepressants? *J Clin Psychiatry* 61:712–721
- Guy W (1976): Early Clinical Drug Evaluation Program (ECDEU), *Assessment Manual of Psychopharmacology*. Rockville, MD, Dept. of Health, Education and Welfare
- Hamilton M (1960): A rating scale for depression. *J Neurology Neurosurgery Psychiatry* 23:56–62
- Haskell DS, DiMascio A, Prusoff B (1975): Rapidity of symptom reduction in depressions treated with amitriptyline. *J Nerv Ment Dis* 160:24–33
- Healy D (1998): Reboxetine, fluoxetine and social functioning as an outcome measure in antidepressant trials: implications. *Primary Care Psychiatry* 4:81–89
- Insel TR (1991): Serotonin in obsessive-compulsive disorder: a causal connection or more monomania about a major monoamine? In Sandler M, Coppen A, Harnett S (eds), *5-Hydroxytryptamine in Psychiatry*. Oxford, Oxford Univ Press, pp 228–247
- Kahn RJ, McNair DM, Lipman RS, Covi L, Rickels K, Downing R, Fisher S, Frankenthaler LM (1979): Imipramine and chlordiazepoxide in depression and anxiety disorders, II, Efficacy in anxious outpatients. *Arch Gen Psychiatry* 43:79–85
- Katz MM, Koslow SH, Berman N, Secunda S, Maas JW, Casper R, Kocsis J, Stokes P (1984): Multivantaged approach to the measurement of behavioral and affect states for clinical and psychobiological research. *Psychol Rep* 55:619–671
- Katz MM, Koslow SH, Maas JW, Frazer A, Bowden CL, Casper R, Croughan J, Kocsis J, Redmond E (1987): The timing, specificity and clinical prediction of tricyclic drug effects in depression. *Psychol Med* 17:297–309
- Katz MM, Wetzler S, Koslow SH, Secunda S (1989): Video methodology in the study of psychopathology and the treatment of depression. *Psychiatric Annals* 19:372–381
- Katz MM, Koslow SH, Maas JW, Frazer A, Kocsis J, Secunda S, Bowden CL, Casper R (1991): Identifying the specific clinical actions of amitriptyline: Interrelationships of behavior, affect, and plasma levels in depression. *Psychol Med* 21:599–611
- Katz MM, Koslow SH, Frazer A (1997): Onset of antidepressant action: Reexamining the structure of depression and multiple drug actions. *Depress Anxiety* 4:257–267
- Katz MM (1998): Need for a new paradigm for the clinical trials of antidepressants. *Neuropsychopharmacology* 19:517–522
- Khan A, Leventhal RM, Khan S, Brown WA (2002): Severity of depression and response to antidepressants and placebo: an analysis of the FOOd and drug Administration database. *J Clin Psychopharmacol* 22:1–6
- Khan A, Warner MA, Brown WA (2000): Symptom reduction

- and suicide risk in patients treated with placebo in antidepressant trials. *Arch Gen Psychiatry* 57:311–317
- Kielholz P, Poldinger W (1968): Die behandlung endogener depressionen mit psychopharmaka. *Dt med Wschr* 93:701–704
- Kuhn R (1958): The treatment of depressive states with G 22355 (imipramine hydrochloride). *Am J Psychiatry* 115:459–464
- Laska EM, Siegel C (1995): Characterizing onset in psychopharmacological clinical trials. *Psychopharmacol Bull* 31:41–44
- Leon AC (2001): Measuring onset of antidepressant action in clinical trials; An overview of definitions and methodology. *J Clin Psychiatry* 62(suppl 4):12–16
- Maas JW, Katz MM, Frazer A, Stokes PE, Swann AC, Dais JM, Casper R, Berman (1991): Current evidence regarding biological hypotheses of depression and accompanying pathophysiological processes: A critique and syntheses of results using clinical and basic research results. *Integrative Psychiatry* 7:155–169
- Miller IW, Keitner GI, Schatzberg AJ, Klein DN, Thase ME, Rush AJ, Markowitz JC, Schlager DS, Kornstein SG, Davis SM, Harrison WM, Keller MB (1998): The treatment of chronic depression, Part 3: Psychosocial functioning before and after treatment with sertraline or imipramine. *J Clin Psychiatry* 59:608–619
- Montgomery S (1995): Are two week trials sufficient to indicate efficacy? *Psychopharmacol Bull* 31:29–35
- Montgomery S, Asberg M (1979): A new depression rating scale designed to be sensitive to change. *Br J Psychiatry* 134:382–389
- Niklson IA, Reimnitz PE, Sennel C (1997): Factors that influence the outcome of placebo-controlled antidepressant clinical trials. *Psychopharmacol Bull* 33:41–51
- Norquist G, Lebowitz B, Hyman S (1999): Expanding the frontier of treatment research. *Prevention and Treatment* 2: Article 0001a (on line)
- Prien RF, Blaine JD, Levin J (1985): Antidepressant drug therapy: The role of the new antidepressants. *Hosp Community Psychiatry* 36:513–516
- Raskin A, Schulterbrandt J, Reatig N, Rice CE (1967): Factors of psychopathology in interview, ward behavior, and self-report ratings of hospitalized depressives. *J Consult Psychol* 31:270–278
- Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger JE, Burns CT (1986): The inventory of depressive symptomatology (IDS): Preliminary findings. *Psychopharmacol Bull* 22:985–990
- Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH (1996): The inventory of depressive symptomatology (IDS): Psychometric properties. *Psychol Med* 26:477–486
- Stahl SM, Shayegan D (2001): Reducing measurement variability in psychopharmacology applying principles of adult education by utilizing multimedia to facilitate changes in rater behavior. *NCDEU Abstracts*, Phoenix, Arizona
- Stassen HH, Delini-Stula A, Angst J (1993): Time course of improvement under antidepressant treatment: A survival analytic approach. *Eur Neuropsychopharmacol* 3:127–135
- Stassen HH, Angst J, Delini-Stula A (1997): Delayed onset of action of antidepressant drugs? Survey of recent results. *Eur Psychiatry* 12:166–176
- Trivedi MH, Rush AJ (1994): Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology* 11:33–43
- Uhlenhuth EH, Matuzas W, Warner TD, Thompson PM (1997): Growing placebo response rate: The problem in recent therapeutic trials? *Psychopharmacol Bull* 33:31–39
- Weissman M, Klerman GL, Paykel ES (1971): Clinical evaluations of hostility in depression. *Am J Psychiatry* 128:41–46
- Weissman M, Bothwell S (1976): Assessment of social adjustment by patient self-report. *Arch Gen Psychiatry* 33:1111–1115