Chapter III

PROPOSED GUIDELINE STATEMENTS

FOR

CLINICAL EVALUATION OF ANTIDEPRESSANT DRUGS

Prepared for the FDA-ACNP Task Force

by the Committee on Anti-Depressant Drug Evaluation

Gerald L. Klerman, M.D., Chairman

Revised April, 1974

INTRODUCTION

These proposals are recommended revisions of the FDA guidelines for testing of anti-depressant psychotropic agents. The initial guidelines, of which these are proposed revisions, were developed in 1970-71 by FDA staff after consultation with representatives from the drug industry and the scientific community and were published unofficially in the April, 1972 issue of NIMH Psychopharmacology Bulletin.

The current proposals were developed by the ACNP-FDA Task Force organized in Spring, 1972 in response to a request from the FDA Division of Neuropharmacological Products.

Guidelines, purposes, descriptions of phases, general methodology and procedures for Early Clinical Pharmacology (Phase I studies) are unchanged from the initial set of guidelines.

PHASE II STUDIES

1.0 Objectives

In Phase II the overall objectives are:

(1) To identify conditions or symptoms which may be therapeutically responsive to the drug.

(2) To estimate the appropriate clinical dosage and duration of effect.

(3) To identify adverse effects.

2.0 Subjects Selection

Early Phase II studies can be conducted with inpatient or outpatient adult males and females who are not pregnant or planning pregnancy in the near future depending on the degree of clinical supervision possible. Other patient categories may be used (children, aged, females who may become pregnant) when proper animal toxicity studies have been conducted and appropriate dosage for these samples is established by cautious and gradual increments. Patients should be continually monitored for safety of the drug, as described in the Phase I outline.

As determined by comprehensive clinical and laboratory evaluations, patients evaluated early in this phase should require no concomitant medication and have no organic diseases that may obscure clinical observations, laboratory tests or interpretations.

In early Phase II studies, heterogeneous samples may be useful to determine range of implications. Later in Phase II, the sample should be as homogeneous as possible, considering the following: age, sex, weight, treatment setting and socio-occupational group.

2.1 Criteria for Selection

Carefully specified criteria for selection of subjects for anti-depressant drug trials are desirable for two reasons:

(1) The greater the homogeneity of the sample, the greater the likelihood of obtaining reliable and valid results in comparing a new compound with placebo or standard drug.

(2) In generalizing the results of clinical trials to therapeutic practice, it is necessary to have an adequate description of the patients' characteristics to provide assurance they are representative of some clinically relevant population.

In evaluation of anti-depressant drugs, major problems arise from the semantic confusions associated with multiple meanings of the term "depression". Depression has different usages in various scientific fields, including physiology, psychology and psychiatry. In clinical psychiatry, the two uses that are most pertinent to drug investigations are depression as a symptom and as a syndrome. As a symptom, depression may occur in association with a variety of medical and psychiatric disorders. As a syndrome, depression may occur in association with other mental, behavioral and psychophysiological manifestations in various, and as yet poorly understood, complex combinations.

Failure to clarify the purpose of the study may lead to difficulties in patient selection, in design and in execution of the study and in unnecessary problems in the interpretation of the data.

2.2 Depression as a Symptom as a Target for Drug Therapy

Depression as a symptom occurs in a wide variety of clinical conditions. Anti-depressant drug treatment may be useful in such conditions. However, research and clinical experience have not clearly demonstrated that drug treatment of depression as a symptom can be generalized across all states.

We also do not have normative data for distinguishing between the normal depressive state and the pathologic symptom. Some preliminary psychometric studies have been undertaken by Beck, Katz, Zung and others, and research efforts are underway using population survey and epidemiologic techniques to generate such norms.

The following recommendations seem appropriate:

(1) Depressive symptoms occur in many transient conditions, such as post partum, following coronary disease, and also as significant accompaniments of many chronic medical ailments of the cardiovascular, rheumatic and arthritic, and gastrointestinal systems. Where a drug is being evaluated for depression in these states, the pharmaceutical firm and the investigators should specify the population, and the criteria for the primary medical diagnoses should be described in detail.

(2) Symptoms required to be present at entrance to the study should be specified in the protocol. Two alternative procedures have been used successfully. They are:

(a) Listing of target symptoms, e.g. depression, crying, agitation, insomnia, etc. with a statement as to their frequency in the sample studies, measure of severity and the duration prior to entrance into the study;

(b) Where established scales are used, pre-treatment level should be established below which patients will not be admitted to the study.

(3) Where anti-depressant drugs are being used for depressive symptoms accompanying other psychiatric conditions, this indication should be specified in the protocol, and the diagnosis criteria identified. In practice, three such populations have been involved in anti-depressant drug trials. These are:

(a) Alcoholics. Care should be taken to exclude patients who are or have only very recently experienced acute withdrawal symptoms such as DT's, tremulousness or seizures.

(b) Schizophrenic or schizo-affective states. The criteria for the diagnosis should be specified.

(c) Aged patients, including those with DNS disease in whom depression is a frequent concommitant.

2.3 The Depressive Syndromes and Drug Evaluation

The most common clinical conditions for which anti-depressant drugs are evaluated are the various depressive syndromes. Syndrome refers to

the temporal coexistence and covariation of related symptoms and behaviors. Patients with a depressive syndrome usually manifest depressed mood (as described below) plus a significant number (4-5) of associated symptoms.

1. Depressed mood characterized by any of the following: sad, low, blue, despondent, hopeless, gloomy
2. Anhedonia - inability to experience pleasure
3. Poor appetite or weight loss
4. Sleep difficulty (insomnia or hypersomnia)
5. Loss of energy; fatigue; lethargy
6. Agitation
7. Retardation
8. Decrease in libido
9. Loss of interest in work and usual activities
10. Feelings of self reproach or guilt
11. Diminished ability to think or concentrate, such as slowed thinking or mixed-up thoughts
12. Thoughts of death and/or suicide attempts
13. Feelings of helplessness and hopelessness
14. Anxiety or tension
15. Bodily complaints

Since these symptoms will occur together at greater than chance, factor analysis and correlational statistics are useful to identify symptom groupings. Depressive syndromes usually involve various complex patterns or configurations. It is important to acknowledge that there is no <u>one</u> depressive syndrome. There is no agreement as to the bases upon which the various depressive syndromes should be identified and separated. Most investigators

currently accept the concept of heterogeneity within the affective disorders and employ various pluralistic or dualistic diagnostic distinctions. The most used pluralistic distinctions are embodied in the official APA-WHO nomenclatures which designate multiple affective states such as manic depressive illness, involutional states, psychoneurotic depression, etc. The dualistic distinctions involve the primary-secondary separation developed by Robins and Guze, the unipolar-bipolar distinction developed by Perris, Leonhard, and Winokur, and the endogenous-neurotic distinction recently revised by Kiloh, Mendels, Klein and others. Older distinctions, which have also been used in drug evaluation, are the retarded-agitated typology and the psychotic-neurotic forms.

In addition, there are statistically derived typologies, which utilize computer programs for assignment of patients to multiple groups. The Overall-Hollister typology has been used in anti-depressant drug studies. Techniques developed by Paykel, by Grinker and associates in Chicago and by Friedman in Philadelphia have not yet been widely applied to drug evaluation.

It is suggested that:

(1) The investigator should specify the classification approach used and the criteria for assignment of patients to designated groups.

(2) Among the assignment and selection techniques shown to have been effective in drug studies are those developed by Raskin for the NIMH collaborative study, by Overall and Hollister for VA studies, by the UK-MRC studies and by Kiloh and Roth for endogenous-reactive types.

(3) Investigators are encouraged to make use of scales which have already been used in drug research. These are described in detail in the reports appendaged. For a number of scales, such as the Raskin, Beck, Wittenborn, MMPI, Hamilton and Zung, information exists as to the range of scores expected in populations, such as outpatient and inpatient and for grades for severity.

(4) The degree of severity of patient's illness should be described using global scales.

(5) The patient's status should be described, i.e., outpatient, day or inpatient and whether the patient is seen in private practice or in group practice, a clinic, or institutional setting.

(6) It is desirable that a variety of socio-demographic and clinical characteristics of the patients be reported. These characteristics include age, sex, racial or ethnic background, social class, previous hospitalization, previous diagnoses of mania or schizophrenia, and previous major therapies, i.e. ECT, phenothiazines, etc.

(7) It is recommended that the criteria for exclusion be identified particularly for patients in the borderline between schizophrenia and the affective diseases.

(8) Problems in design of studies and in selection of patients with anxiety and depression will be discussed elsewhere in these guidelines.

3.0 Sample Size

In early Phase II studies, sample size may vary since actual numbers depend on the problems to be investigated and on the magnitude of difference between the treatment and control groups that is to be expected or is actually observed. There should be sufficient numbers of subjects to assure a reasonable likelihood of demonstrating differences if they exist.

In later Phase II or Phase III studies, a minimum of twenty patients per treatment group are necessary if drug-placebo or drug-drug differences are to be demonstrated.

4.0 Setting

In early Phase II studies, inpatient setting should be used when feasible. If other settings, e.g. outpatient clinic, day hospital, private clinic or private office are necessary, these should be described and justified.

5.0 Investigators

The investigators should be experienced in evaluating psychiatric drugs and in the conduct of clinical trials; they should have ready access to the appropriate population group for whom the drug may be indicated.

6.0 Design

Patients should be selected to provide an unbiased sample of the population of interest and should be assigned to treatments at random. Pre-drug severity of the disorders to be measured should be recorded and included as part of the design.

6.1 Drug Free Period

When safe and feasible, each study subject should have a drug free period for several days prior to receiving the study medication. The number of days would depend upon the prior medication received by the subject and its duration.

There are two reasons for this procedure. One, rapid remitters and placebo responders can be detected and eliminated from the study, thus maximizing the likelihood of establishing drug effect. Second, during the "wash out" period, patients who have been taking drugs with potential for dependence or withdrawal will manifest these behaviors.

If patients' symptoms decrease to level below criteria for entrance into the study during this "wash out" period, they should not continue in the study.

6.2 Uncontrolled Trials

In early Phase II, several open or uncontrolled trials may be desirable to allow investigators sufficient flexibility to explore possible aspects of a new drug's activity and to allow for the determination of an appropriate dosage range for use in double-blind studies. It must be kept in mind that information obtained from these early open studies can only form hypotheses which must then be tested in controlled studies. The open studies may be of small sample size; however, it may be desirable for validated clinical measures and selection of patient samples to be consistent between investigators in order to facilitate the interpretation of results.

6.3 Controlled Trials

Hypotheses evolved from open studies can be confirmed or refuted only by controlled double-blind studies. In at least some late Phase II studies the investigational new drug should be compared to a matching placebo control to establish its efficacy. Other studies may include only an active treatment control or both.

Parallel groups, cross-over, intensive and other designs may be used. The planning of this and other aspects of these studies should, whenever feasible, involve extensive consultation with a biostatistician. There should be full awareness of the advantages, disadvantages and criteria for validity regarding each possible design before selecting one. Comparisons between two active drugs usually require much larger sample sizes than active drug-placebo comparisons.

Packaging and coding of medications should be performed on an individual basis rather than on a treatment group basis. Other psychoactive drugs are to be avoided. If other drugs are used, this should be carefully documented.

7.0 Dosage

7.1 Open or Uncontrolled Studies

After selection of initial dosage based on all previous data (including pharmacokinetic), dosage in open trials is usually increased until a satisfactory therapeutic response is observed. If adverse effects are a significant problem, further increases may be precluded and dosage reduction or discontinuation indicated.

7.2 Double-blind, Controlled Studies

Dosage may be fixed; however, because of individual metabolism and tolerance, a flexible dosage may be necessary.

A specified range may be used (as determined in earlier trials) within which adjustments are made individually according to specified clinical criteria.

The mode of administration, range, schedule of administration and criteria for dosage adjustment should be stated in each protocol.

8.0 Duration of Trial

The duration of individual clinical studies in Phase II may vary from days to weeks depending on their purpose and the nature of the drug. The therapeutic activity of anti-depressant agents usually can be established in trials of approximately four weeks duration. However, to allow for a further assessment of safety beyond that obtained in Phase II, at least one of the first several Phase II studies should be continued for six weeks with appropriate laboratory monitoring if preliminary data have indicated satisfactory support of efficacy.

9.0 Assessment

Physical exams and clinical laboratory tests are basically the same as those for Phase I. Appropriate and validated rating scales should be used in addition to global assessment; provisions should be made for recording symptom emergency which may represent side effects or possible new uses for a drug. The reliability of the ratings may require more than one rater.

Baseline observations should be carried out in all patients immediately before their initiation into the study. The frequency of follow-up determinations may then vary from days to weeks.

All adverse reactions, reported or observed are to be part of the record. Include the dose of the drug at the time of appearance of the side effect, the duration of the side effect, severity, judgment as to whether the drug is responsible for the side effect, method of treating the side effect, and the results of the treatment. If the side effect was severe enough to discontinue the subject from the study, this should be stated.

9.1 Techniques for Assessment of Change

There are a number of techniques which can be applied in assessment of change with anti-depressant drugs. No one technique in itself is considered sufficient. Not all techniques are required. Varying combinations may be employed depending on the phase of investigation and the types of patients.

These techniques include:

(1) Global measures. The NIMH and VA studies have used 5-7 point scales. The Menninger Health-Illness Scale may also be used. Categories of improvement (marked-moderate-some-none-worse) are widely used.

(2) Clinical Interview Scales. The scales developed by Hamilton, Levi, Overall, Spitzer, Wittenborn and others are of demonstrated reliability, validity and sensitivity to drug effects.

(3) Self report techniques. These have gained wide acceptance in outpatient samples. Among those frequently used are the Zung, Beck, the MMPI, the Symptom Check List (SCL), the Clyde Mood Scale (CMS), and the Adjective Check List. Their use has been reviewed in detail by McNair, and his paper should be consulted for those scales which have proven useful.

(4) For inpatients and day patients, direct observational approaches to behavior rated by nurses or other personnel have been used. The most common scales are those of Lorr, the Burdock, the NOSIE, and the Grosser-Wechsler scales.

(5) Social adjustment. Assessment of social adjustment seems most appropriate for drug maintainance trials, where patients are followed into the community after discharge from the hospital. In these studies, drug therapy attempts at demonstrating efficacy in the prevention of relapse and recurrence is prophylactic or maintenance therapy. This need occurs in long term trials with Lithium or tricyclics. These techniques are reviewed in the paper prepared by Weissman.

(6) Psychological techniques, such as projective methods and tests of intelligence, have not been used widely as measures of efficacy. They are useful as associated data, and further research may demonstrate their validity.

(7) Verbal sample techniques and content analysis, such as those developed by Gottschalk and associates, have proven sensitive to drug response.

(8) Psychophysiological measures, including EEG, pulse rate, blood pressure, EMG, GSR, pulse volume, do not in themselves provide evidence of efficacy but may be useful as associated indices of change. However, pulse, BP and weight are valuable clinical indicators and suggested for studies, even though they may not be criteria of efficacy. Specialized psychophysiological methods, such as quantitative EEG analysis, while still in the investigational stage, may provide valuable data.

Other techniques, including rating scales, developed by the investigator may be employed providing evidence for their reliability and validity is available.

Efforts should be made to establish agreement on the use of diagnostic and descriptive terms as well as the handling of assessment instruments. This may be accomplished by investigator meetings or in depth discussion by the monitor with each investigator.

10.0 Interpretation

In addition to clinical considerations, the interpretation of studies in this difficult area may involve extensive statistical analysis which often cannot be predetermined. Rather than interpreting studies only individually, it is also important to consider all information available concerning a drug, including preclinical studies, before making final positive or negative conclusions regarding it.

Insofar as possible, it is desirable to demonstrate efficacy within the

the unit study. However, it may be difficult to obtain enough patients in a unit study to achieve strong statistical evidence of efficacy therein, particularly when two active medications are involved. This may result simply from a scarcity of patients, or from particularly strong placebo effects in the type of patient and/or setting involved. Under such circumstances, it may be necessary to pool data from several studies to provide convincing statistical evidence of efficacy. This decision should be made in research protocol planning. The pooling of data from different investigators can be of definite value but requires special consideration.

PHASE III STUDIES

By the middle or end of Phase II, sufficient information should be available to formulate hypotheses as to types of patients and their clinical conditions which may respond to the investigational drug. For testing these hypotheses, controlled trials are necessary. This is the major purpose of Phase III studies.

1.0 Specific Objectives

(1) Therapeutic Studies. Extension of comparative controlled studies are used to fully confirm the drug's basic anti-depressant activity in heterogeneous patient populations and to provide more specific information about symptoms and patient types in which the drug is especially effective. Placebo controlled trials are necessary as are comparisons with standard drugs of established efficacy for the clinical condition.

(2) Long Term Safety Studies. To establish the safety of a new anti-depressant agent when given daily for 3 to 6 months, long term safety studies are undertaken with particular regard to the nature, incidence and control of side effects.

2.0 Patient Selection

Similar considerations apply as in Phase II. In this phase, patients with the diagnosis of primary depressive syndrome of some form are usually selected. Populations with other than primary depressive disorders, such

as depression with schizophrenia, or in association with anxiety or with medical conditions may be studied in separate trials.

A greater variety of populations differing as to age, sex, diagnostic categories, social class, treatment setting, previous treatment, etc. may be studied. Within each study (or subgroup in studies of sufficient size) patients should be selected to be as homogeneous as possible regarding the above variables. In any case, full reporting of patient characteristics is necessary to allow for adequate interpretation of results.

Exclusions should be stated. Exclusion of placebo responders may strengthen a trial. The placebo response for a particular patient population should, at least, be known to the investigator.

Females of child-bearing age may be included if results of animal reproductive and teratologic studies are satisfactory.

3.0 Sample Size

While it may be possible to demonstrate drug effects in samples as small as twenty patients per treatment group, experience has shown that samples in the range of thirty to fifty patients in each treatment group provide greater assurance.

These considerations are discussed in greater detail in the chapter prepared by Wittenborn.

4.0 Setting

A number of settings may be used, e.g. inpatients, outpatients, private practice.

5.0 Investigators

Because the extension of claims into other areas involves a variety of types of investigators, it is important to consider the investigator's capability and experience in evaluating and working with patients with depression, his provisions for needed safety precautions, his access to laboratory facilities with suitable controls, and the appropriateness of his clinical setting to allow for valid drug evaluation.

6.0 Design

Of primary importance during Phase III of a drug's evaluation are controlled studies designed to fully confirm the drug's basic antidepressant efficacy. The design guidelines are generally the same as those discussed in Phase II. However, adjustments may be made in controls, duration of study, dosage and design which do not interfere with validity to accommodate greater variations in purpose of studies, settings, investigators and subjects as discussed under the respective headings in Phase II.

There is wide agreement that a number (preferably three to five) of studies in Phase III compare the new compound with a placebo and with an established drug of demonstrated efficacy in similar pharmacological class (tricyclic, MAO inhibitor, psychomotor stimulant). These studies should be double-blind and utilize standard methods of assessment of change.

Crossover designs to control bias have major inherent difficulties because of carryover effects, both pharmacological and psychological, particularly in depressive syndromes (spontaneous remissions, self-limiting, response to attention, etc.)

When it is concluded that the drug's basic anti-depressant efficacy has been clearly established by controlled studies, consideration may be given to undertaking further studies on an open trial basis with new populations. These, of course, carry with them the inherent risk, due to lack of a control group for comparison, of encountering difficulties in interpretation of unexpected findings. However, such findings, as stated previously, can lead to forming hypotheses which must be confirmed or refuted by reviewing already completed or establishing further controlled studies. In providng further experience with the drug (often under conditions of usual medical practice), these studies can be important in providing corroborative support of efficacy demonstrated by well controlled studies and in adding valuable data regarding safety of the new drug. This is particularly true when a number of investigators working independently obtain similar findings.

Patients may also be evaluted in studies related to other than the major anti-depressant claim. For each of these other areas, a number of double-blind studies should be sufficient. Prior to these it may be necessary to carry out several open studies to familiarize the investigator with the drug's activity and appropriate dosage range in a particular therapeutic situation or special population. In these populations carefully edited (i.e. they need not include all variables of other studies) clinical pharmacology and therapeutic trials may be carried out to establish tolerance, efficacy and safety.

7.0 Long Term Safety Studies

Long term safety studies may be on an open trial basis or may be of a controlled parallel groups design.

Data regarding long term safety may also be obtained from a number of studies rather than from a single formally structured one. For example, provision may be made for patients in therapeutic trials to continue on the drug if it is indicated. Special attention will need to be given to children, elderly patients and women who are of child bearing potential.

8.0 Dosage

It is necessary to establish in dose response studies the dosage that gives adequate clinical response with minimum of side effects.

In long term safety studies dosage should be of at least the level expected for eventual general therapeutic use. Allowance for adjustments to age, sex and individual tolerance, of course, must be provided.

9.0 Duration

The duration of therapeutic trials (Phase III) may vary as in Phase II and, particularly regarding open trials, may often be longer. After baseline determinations in all cases, follow-up evaluations usually become less frequent as evolving data permits.

The duration of long term safety studies is usually three to six months. Baseline observations should be carried out in all cases and follow-up evaluations are usually at least monthly.

10.0 Assessment

This is generally similar to that outlined for Phase II. Valid adjustments are permissable as indicated. Laboratory evaluations are usually not necessary for as many patients as in Phase II on the basis of accumulated data.

11.0 Interpretation

Since most Phase III studies will involve multiple trials, consideration must be given to problems of statistical analysis and interpretation within each trial, among various trials, and when data from a number of trials are grouped or pooled. These issues are reviewed in greater detail in the chapter written by Wittenborn for this Task Force and in the section on Documentation and Interpretation in Levine, Schiele, and Bouthilet, Principles and Problems in Establishing the Efficacy of Psychotropic Drugs, 1971, Washington, D.C. U.S.G.P.O.

MAINTENANCE STUDIES AND LONG TERM TRIALS

Currently there is considerable interest in long term trials to establish the value of maintenance anti-depressant therapy in prevention of relapse or recurrence. Whereas the treatment goal in acute episodes is the reduction of symptoms and the patient's return to a premorbid state, in long term maintenance therapies, the goals are more ambitious: prevention of relapse or recurrence after an initial episode, facilitation of the patient's social and vocational adaptation, relief from minor symptoms, enhancement of personal adjustment, and life satisfaction.

Research methodology in these studies is still in the developmental stage and there are many unsolved design and statistical problems.

A major problem is the identification of patients likely to suffer from multiple recurrences. While a significant number of patients with acute depressions are likely to have multiple recurrences, it is difficult, on the basis of presenting clinical symptoms, to predict which patients experiencing acute depressions are likely to have recurrences. The frequency of recurrence in a large population, as well as the duration of intervals between relapse and recurrences, is highly variable.

Given this lack of information, investigators have employed different strategies to select samples for long term maintenance studies. Most investigators advocate selecting patients who have had two or more well-defined episodes and/or hospitalizations. Others have proposed selecting patients with salient symptoms or characteristics such as familial history, manic elations, clear-cut retardation, or well-defined periodicity.

In evaluating the efficacy of treatment, the criteria of effectiveness are different from that used in studies of acute treatment. Reduction of mortality, especially from suicide, is a major criterion. Another is prevention or reduction of rehospitalization rates. These effects are more easily identified, but occur less frequently than minor relapses or recurrences, which may not be of sufficient intensity to require hospitalization. For the detection of such symptomatic change, rating scales and systematic clinical observation are needed.

Moreover, for evaluating criteria in areas such as familial and marital functioning, social effectiveness, and vocational performance, new techniques are needed in addition to the established rating scales which assess psychopathology and mood.

Design problems for long term maintenance studies are considerable. In the absence of established data as to the predictors of relapse and the frequency of rehospitalization and recurrence, a concurrent control group, receiving no treatment or placebo, is highly desirable.

There are significant statistical problems, particularly the need for statistical techniques for repeated measures. Various forms of trend analysis or stochastic models may assume an important role.

A related problem deals with attrition. The longer a trial continues, the greater the likelihood that patients will drop out; the sample finishing the trial will not be identical to that beginning the trial. Life-table methods and other statistical techniques, derived from public health morbidity and mortality studies, may have applicability in this area.