The Commonwealth of Massachusetts
Department of Mental Health

GERALD L. KLERMAN, M.D.
SUPERINTENDENT

ERICH LINDEMANN MENTAL HEALTH CENTER
GOVERNMENT CENTER
BOSTON, MASSACHUSETTS 02114

6 February 1973

To: J.R. Wittenborn, Ph.D.

From: Gerald L. Klerman, M.D.

Subject: SOME GENERAL CONSIDERATION REGARDING THE PREPARATION OF GUIDELINES FOR EVALUATING ANTI-DEPRESSANT AND ANTI-ANXIETY DRUGS

The purpose of this memo is to record some of my thoughts about our efforts to prepare new guidelines for the evaluation of anti-depressant and anti-anxiety drugs. These efforts are part of the project undertaken per contract between the ACNP and FDA.

Considerable discussion has been generated by the ACNP's undertaking of this project. Some have expressed doubt as to the wisdom of guidelines. Others have questioned the need for revision of the anti-depressant and anti-anxiety drug evaluation guidelines, since the initial guidelines have just recently been published. This criticism is ironic, since at the time the guidelines were first initiated by the FDA, concern was expressed by many investigators les the guidelines become rigid and not open to revision. Now there is some concern that the guidelines are being revised too rapidly.

The impetus for our current efforts arises from disatisfaction with the initial version of the guidelines, they do not provide sufficient detail and specificity. The felt need for more specific guidelines arises from the large number of NDA applications for various classes of drugs proposed to be used for the treatment of anxiety and depression, classes of drugs whose initial NDA approval were for other clinical indications. With the awareness that the largest number of patients receiving prescription psychotropic drugs are outpatients with neurotic affective states, particularly anxiety and depression and their combination, the need for more detailed and specific guidelines was felt by the FDA to be of high priority.

THE PLACE OF GUIDELINES IN A COMPREHENSIVE FDA PROGRAM OF IMPROVING DRUG EVALUATION

The revision of guidelines is but one component, although a necessary one, of a comprehensive effort now being undertaken by the FDA to improve the quality of evaluation of new drugs and the process of their NDA review by FDA.

As I understand FDA intentions, this comprehensive program involves four components which include:

(1) Guidelines, revised and updated as scientific standards evolve;

(2) Increasing use by FDA of outside consultants to evaluate IND and NDA applications along with the FDA staff. As part of this increasing use of outside consultants, advisory committees such as the Advisory Committee on Neuropharmacology, are being established on a standing basis;

(3) Improvement in the presentation of data accompanying IND and NDA applications. There is general agreement that the current format in which data are presented to FDA by the pharmaceutical firms is grossly inadequate. Too much trivial data are provided, and the reviewer is overwhelmed by the minutia and disorganization. Moreover, the format is often inconsistant and lacks standardization. Attempts to provide standardized format are underway, possibly to develop computerized data presentations.

(4) Periodic conferences between the pharamceutical firm, investigators, FDA staff, and FDA consultants. These conferences are expect to take place at significant points in the review process, particularly the beginning of Phase 1 and Phase 2 as well as prio to the NDA. It is hoped that these conferences will provide a mechanism for better exchange between the parties concerned and to increase likelihood of adequate evaluations when the drug application reaches the final NDA review.

RATIONALE FOR GUIDELINES

The need for guidelines seems to follow inevitably from the 1962 Kefauver-Harris amendments to the FDA statutes. In these amendments, Congress added efficacy to the previously mandated criteria of safety and purity. If efficacy is to be a criteria, then the evidential procedures and processes by which efficacy is to be judged, need to be established to provide due process for all parties concerned.

In this respect, it is significant that the development of the controlled clinical trial as a specific research methodology seems to coincide with efforts by the FDA in the United States and by similar groups in other nations to improve the quality of drug evaluations.

The purpose of guidelines is to delineate procedures and processes which will better specify the nature of evidence to judge efficacy and safety. Considerations such as selection of patients, research design, drug dosage and administration schedules, modes of assessment and statistical analyses.

The fear has arisen that these guidelines could result in rigid restrictions on the advance of clinical pharmacology. This is of concern to many investigators, since the guidelines bear the approval of a governmental agency. There is general agreement, however, that such guidelines should reflect the most advanced scientific knowledge and also should be structured in such a way as to provide a stimulus for new research on methodology and clinical pharmacology. It is generally agreed that guidelines should be updated periodically to reflect changing scientific standards and new knowlege.

AREAS OF NEEDED RESEARCH

The development of these guidelines in psychopharmacology has highlighted a number of areas in which the available scientific data are inadequate to provide the degree of specification felt desirable. As regards anti-depressant drug evaluation, it is possible to identify four areas . These are:

(1) Better psychometric criteria for psychopathology. We lack validated quantitative norms for differentiating normal depressive mood from abnormal symptomatic states as well as for delineating the constellation of symptoms and behaviors that constitute necessary and sufficient criteria for assigning patients to the various affective syndromes. Potential data relevent to these criteria and norms exists in the data files of the Ecdeu project and with individual investigators, but extensive secondary analyses will be required and in other instances, new data are required. In the final report from this Committee, I will recommend that NIMH and FDA cooperate to identify such data sources and the procedures for their exploitation.

(2) Better criteria for clinically meaningful change. As will be described below, most recent efforts have been to demonstrate drug-placebo differences on specific scales. The clinical relevence of these changes is often unclear. Efforts are needed to define clinical end points, i.e., symptom free state or remission. In addition, return to social functioning or other measures of personal functioning should be explored as criterion measures.

(3) Further experience with techniques for establishing and training of observers for inter-judge agreement are called for.

(4) Most change assessments have used rating scales and self report inventories. Non behavioral criteria are needed such as EEG, psychophysiological measures, such as muscle tension, blood flow, etc., or biochemical assays. At the present time, these measures are in the research stage, but hopefully they may be sufficiently validated to be useful in clinical evaluation.

CONCLUSIONS

The development of guidelines involves a conference of public need with professional interest. As new drugs are developed, there will be increased pressure for public specification of the criteria by which they will be judged safe and efficacious. Moreover, as third party payment programs increase their coverage of psychiatric illnesses, pressures can be expected to grow for determining as to whether or not the treatments for which payments are requested are effective, safe, and of relative low cost. It is in the best interest of the profession as well as meeting the needs of the public for scientific investigators and clinicians to participate actively in the development of these standards and criteria.

The Commonwealth of Massachusetts
Department of Mental Health

GERALD L. KLERMAN, M.D.
SUPERINTENDENT

ERICH LINDEMANN MENTAL HEALTH CENTER
GOVERNMENT CENTER
BOSTON, MASSACHUSETTS 02114

6 February 1973

Memo

To: Richard Wittenborn, PhD.

From: Gerald L. Klerman, M.D.

Re: EVALUATING ANTI-DEPRESSANT DRUGS

I PURPOSES

This is a preliminary draft of proposed guidelines for evaluating anti-depressant drugs. The final version of the guidelines will need to incorporate this material into the initial guidelines formulated in 1970.

This version is intended for discussion among members of the FDA-ACNP project, the FDA Advisory Committee on Neuropharmacology, and selected persons active in the field, including representatives of the pharmaceutical industry, clinicians, and investigators.

II SELECTION OF PATIENTS

It is agreed that the new guidelines should specify procedures and criteria for selection of patients for anti-depressant drug evaluation. Improved selection is desirable for two reasons: first, the greater the homogeneity of the sample, the greater the likelihood of obtaining reliable and valid results when comparing a new drug with a standard and/or a placebo; and, second, in generalizing the results of research trials to clinical practice, it is necessary to have an adequate description of the clinical and social characteristics of the patients and the assurance that they are representative of some clinically relevant population.

In evaluation of treatments for depression, major problems arise from the semantic confusion due to the multiple meanings of depression. As Lehmann, Mendels, and Klerman have pointed out, depression has different meanings in various scientific field, such as physiology, psychology, and psychiatry. In clinical psychiatry, the two uses of the term that are most relevant to evaluating drugs are depression as a symptom and depression as a syndrome. The distinctions between these two indications for anti-depressant drug treatment have major consequences for drug evaluation and for labeling and clinical prescription.

As a symptom, depression is a pathological extension of the normal mood of depression, probably best called sadness. Some investigators advocate that the depressive symptom be called dysphoria to distinguish it from the normal mood. Dysphoria occurs in association with a wide variety of clinical states, psychiatric and non-psychiatric. In this sense, the use of a drug for treatment of dysphoria represents an application to the target symptom approach. The distinction between symptom and syndrome is similar, but not identical, to the distinction between state and trait in psychology. This distinction is also actively discussed with respect to anxiety, where anxiety can refer to an emotional state and to the trait anxiety proneness.

Any studies re this? Can we obtain data easily?

Failure to specify the purpose for which the drug is being investigated may lead to difficulties in interpretation.

Depression as a Symptom and as a Target for Drug Therapy

Depression as a symptom, or dysphoria, 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 dysphoria can be generalized across all states. It does not bear the same relationship to drug response as, for example, does pain. In evaluating analgesia, it is generally accepted that pain is similar in different clinical states and that over a wide variety of conditions, drug responses can be generalized, i.e., headache, arthritis, dental pain, etc. There are, however, limits to this concept in analgesia. The general drug-pain response may be reasonable because of the established neurophysiological system of pain receptors and transmission tracts. We do not have comparable knowledge of the CNS regulation of affective states and associated neuroanatomical and neurochemical mechanisms.

Our study with Normals & impairing

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

yes we do

The following recommendations seem appropriate:

(1) Depressive symptoms occur in many conditions as a transient condition, such as post partum, following abortion, 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 medical diagnoses should be discarded.

How about loss reaction to a death, etc.

~~How about~~

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

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

(b) Where established scales are used, some 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 dysphoria symptoms accompanying other psychiatric conditions, this aim should be specified, and the diagnostic groups 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 recently experienced acute withdrawal symptoms such as DT's, tremulousness, or seizures.

(b) Schizo-affective or schizophrenic states; the criteria for the diagnosis should be specified;

(c) Aged patients including those with CNS disease in whom depression is a frequent concommitant:

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 of related symptoms and behaviors. Since the symptoms will occur together at greater than chance, factor analysis and correlational statistics are useful. Syndromes identify various complex patterns or configurations. At the present time, it is important to acknowledge that there is no one 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 distinctions. The most used pluralistic schemata 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 commonly in use include the primary-secondary affective disorder distinction developed by Robins and Guze, the unipolar-bipolar distinction developed by Perris, Leonhard, and Winokur, and the endogenous-neurotic distinction recently explicated by Kiloh, Mendels, and others. Older distinctions, which have also been used in drug evaluation, are retarded-agitated and the psychotic-neurotic forms.

Many available ways of looking at diagnosis

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

We recommend that:

(1) The investigator should specify the schema for classification and the criteria for assignment of patients to specified 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, Overall and Hollister for VA studies, and criteria in the UK-MRC studies and in Kiloh and Roth criteria for endogenous-reactive types.

(3) Investigators should be encouraged to make use of those 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, Zung, tentative information exists as to the range of scores expected in populations such as outpatient and inpatient and for grades of severity.

(4) The degree of severity should be described using global scales.

(5) Institutional 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 other characteristics of patients be reported (see Raskin paper). These characteristics include age, sex, racial or ethnic background, social class, previous hospitalization, previous diagnoses of mania and schizophrenia, and previous major therapies, i.e., ECT.

(7) It is recommended that the criteria for exclusion be identified, particularly in the borderline between schizophrenia and the affective diseases (see NIMH criteria and Feighner criteria). Problems with the overlap with anxiety will be discussed below.

III ASSESSMENT OF CHANGE

There are seven techniques which can be applied in the assessment of change. These can be employed in different combinations depending on the stage of investigation (Phase 2 or Phase 3), the status of the patients (inpatient or outpatient), and the design of study. These seven techniques are:

(1) Global measures of change; NIMH and VA studies employ 5-7 point scales. The Menninger Health illness scale, or categories of improvement which may be developed by the individual investigators, have also been used.

(2) Psychiatric interview scales such as those developed by Hamilton, Lehmann, Lorr, Overall, Wittenborn and others; these have been reviewed in detail in the report by DiMascio.

(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 Lubin 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. These have been reviewed by DiMascio.

(5) Social adjustment; assessment of social adjustment seems most appropriate for drug maintenance 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 prophalactic 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 utility as measures of efficacy.

Yes they have in research studies

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

IV CLINICALLY MEANINGFUL CHANGE

(1) There is general agreement that the most important method for generating clinically meaningful information as to change is the comparison of a new compound against a standard compound and a placebo. Among the standard compounds for anti-depressant drugs are the members of the tricyclic class of which imipramine and amitryptaline are the most used. Other standard compounds include members of the MAO inhibitor class. Where indicated, an established phenothiazine such as thioridizine may be employed as standard drug.

% change as used with Raskin 3 point

(2) Another technique for clinically meaningful change involves clinical judgment by clinicians as to the degree of improvement on a severity scale (normative) and degree of individual change since start of trial (ipsitive).

(3) Equivalence scores of standardized scales corresponding to the degree of improvement are useful. Information on this approach has been reported for Zung and Beck scales.

(4) Criteria including social behavior status such as change from inpatient to outpatient. It is recognized that such criteria are subject to influence by multiple factors other than psychopathology or drug factors such as administrative practices, family tolerance, economic conditions.

(5) Other suggested areas, which have not been widely used in drug trials include the following:

(a) At start of investigation, attempt to specify the magnitude of change desired. Having done so, it is possible to choose confidence limits and to project the sample size likely to be required to achieve these goals.

(b) Where clearly defined endpoints can be specified, i.e. discharge from hospitals, return to pre-morbid level functioning, absence of symptoms, useful criteria are the number and percentage of patients in each treatment group meeting the criteria at a specified point in time.

(c) Also to be explored are measures of patient acceptance and tolerance, such as the patients willingness and desire to continue taking the drug or the extent to which the drug may interfere with the patients normal activities or the patient's judgment as to how close he or she is to normal functioning or to being well.

V TRAINING RATERS TO IMPROVE RELIABILITY

There is little that I can add to the statements prepared by Katz and by the Uhlenhuth committee.

VI CONCLUSION

This memo has focused on the four areas identified in the contract between FDA and ACNP. These are preliminary recommendations. They should be reviewed by the various committees and by outside consultants. I assume the final document will include as appendices edited versions of the review papers prepared by the task force members in the summer and fall of 1972.