CENTER FOR DRUG EVALUATION AND RESEARCH

Guidance for Industry

The FDA published Good Guidance Practices in February 1997. This guidance was developed and issued prior to that date.

Additional copies are available from:
Office of Training and Communications
Division of Communications Management
Drug Information Branch, HFD-210
5600 Fishers Lane
Rockville, MD 20857

(Tel) 301-827-4573
(Internet) http://www.fda.gov/cder/guidance/index.htm

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, FOOD AND DRUG ADMINISTRATION

guidelines for the clinical evaluation of

Psychoactive Drugs in Infants and Children

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Food and Drug Administration

HEW(FDA) 79-3055

GUIDELINES FOR THE CLINICAL EVALUATION OF PSYCHOACTIVE DRUGS IN INFANTS AND CHILDREN

JULY 1979

This publication may be reproduced and distributed without permission of the Food and Drug Administration

Comments on the contents of this publication are invited and should be addressed to the following office:

Group Leader for Psychopharmacological Drugs
Division of Neuropharmacological Drug Products
Bureau of Drugs
Food and Drug Administration
5600 Fishers Lane
Rockville, Maryland 20857

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Stock No. 017-012-00281-1

ABSTRACT

The Food and Drug Administration, with the assistance of its scientific Advisory Committees and other outside consultants, the American Academy of Pediatrics' Committee on Drugs, and consultants to the Pharmaceutical Manufacturers' Association has developed guidelines for the clinical evaluation of new drugs. These guidelines present acceptable current approaches to the study of investigational drugs in man, and pertain to Phases I through III of the investigation. They represent generally accepted principles for arriving at valid conclusions concerning safety and effectiveness of new drugs, as well as the views of outstanding experts concerning appropriate methods of study of specific classes of drugs.

The FDA welcomes comments on the guidelines, and expects to keep them current by review and update at approximately two-year intervals.

FOREWORD

The purpose of these guidelines is to present acceptable current approaches to the study of investigational drugs in man. These guidelines contain both generalities and specifics and were developed from experience with available drugs. It is anticipated that with the passage of time these guidelines will require revision. In order to keep them current a re-review will be performed approximately every 18 to 24 months.

These guidelines are not to be interpreted as mandatory requirements by the FDA to allow continuation of clinical trials with investigational drugs or to obtain approval of a new drug for marketing. These guidelines, in part, contain recommendations for clinical studies which are recognized as desirable approaches to be used in arriving at conclusions concerning safety and effectiveness of new drugs; and in the other part they consist of the views of outstanding experts in the field as to what constitutes appropriate methods of study of specific classes of drugs. In some cases other methods may be equally applicable or newer methods may be preferable, and for certain entirely new entities it is possible that the guidelines may be only minimally applicable.

Under FDA regulations (21 CFR 10.90(b)) all clinical guidelines constitute advisory opinions on an acceptable approach to meeting regulatory requirements, and research begun in good faith under such guidelines will be accepted by the Agency for review purposes unless this guideline (or the relevant portion of it) has been formally rescinded for valid health reasons. This does not imply that results obtained in studies conducted under these guidelines will necessarily result in the approval of an application or that the studies suggested will produce the total clinical information required for approval of a particular drug.

Many of the clinical guidelines have been developed largely, or entirely, by FDA's Advisory Committees and consultants. Others were originally developed by intramural committees and consultants of FDA and of the Pharmaceutical Manufacturers Association; in these cases the guidelines were reviewed and revised, as appropriate, by FDA's Advisory Committees.

The general guidelines for the evaluation of drugs in infants and children and most of those for study of various drug classes in children were developed by the Committee on Drugs of the American Academy of Pediatrics (AAP). Some of the pediatric guidelines for specific classes were written by FDA's Advisory Committees. There was cross review and comment on the pediatric guidelines by both the Committee on Drugs of the AAP and FDA's Advisory Committees.

The Bureau of Drugs of the FDA wishes to thank the many individuals who devoted so much time and effort to the development of these guidelines.

J. Richard Crout, M.D.
Director
Bureau of Drugs

Marion J. Finkel, M.D.
Associate Director for
New Drug Evaluation
Bureau of Drugs

GUIDELINES FOR EVALUATION OF PSYCHOACTIVE AGENTS IN INFANTS AND CHILDREN*

INTRODUCTION

General Considerations for the Clinical Evaluation of Drugs should be consulted.

These guidelines present general considerations for the clinical evaluation of psychoactive drugs to be used in the pediatric population. A paper entitled "General Principles for Psychoactive Drug Studies" by J.R. Wittenborn is attached as Appendix I and has been referenced extensively in the present document because it deals with many important and relevant issues that must be addressed in studying children as well.

These guidelines are intended to help those who design and conduct investigations of psychopharmacologic agents in children. They will be utilized in the review of IND protocols designed to conduct such studies, and in the evaluation of safety and efficacy of psychoactive substances claimed in new drug applications. These guidelines present a somewhat idealized set of criteria. It is, of course, recognized that individual studies may often not be able to meet every criterion considered here. However, such studies are seldom complete or sufficient by themselves. Some reasonable approximation to the criteria set forth in these guidelines can often be achieved by a series of investigations, each carefully attending to a somewhat different set of considerations. Some requirements, such as early demonstration of safety before major clinical studies are undertaken, are relatively fixed, others may be more flexible. There are several problems unique to the pediatric age group which have influenced the recommendations herein concerning the design, conduct of the study and phases through which preparation should proceed, whatever the particular therapeutic area under consideration. Recommendations in the present guidelines are intended to apply to the whole field of psychoactive drugs in chidren, with the expectation that additional guidelines will eventually be needed for specific therapeutic areas.

PRECLINICAL STUDIES

Preclinical testing in animals for pharmacologic activity and toxicology precede any application for investigational new drug (IND) status according to currently recognized FDA Guidelines. Drugs for use during pregnancy and in the pediatric age groups require additional preclinical testing as specified in the "General Considerations for the Clinical Evaluation of Drugs of Infants and Children". This testing may entail the development of techniques to assess, preclinically and clinically, pharmacokinetics (absorption, distribution, metabolism, and excretion), newborn: adult LD50 ratios, mutagenesis, teratogenesis, carcinogenesis, growth, development, sexual maturation, cognitive skills, psychologic development, and reproductive capacity.

*These guidelines were prepared by a Pediatric Subpanel of the Psychopharmacological Agents Advisory Committee. This subpanel consisted of Bonnie Camp, M.D., Ph.D., Donald Robinson, M.D. and Robert Reichler, M.D. Other members of the Pediatric Subcommittee who participated in writing these guidelines are: Rachel Gittelman, Ph.D., Ronald Lipman, Ph.D., Gabrielle Weiss, M.D., Robert Moore, M.D., Lee Robins, Ph.D., Judith Rapoport, M.D., Robert Sprague, Ph.D., Albert DiMascio, Ph.D. and Keith Connors, Ph.D.

Since psychoactive drugs may be administered to children over months to years, special investigations are indicated in preclinical testing. Thus, assessment in animal species should be carried out for possible drug effects on growth, development, pubescence and reproduction. Evaluation should be carried out in two or more appropriate animal species selected on the basis of the comparability of their stages of development to stages of human growth and development. Drug administration should be timed in such a way that effects on selected stages of development can be assessed, particularly periods of rapid growth and development. If serious aberrations of growth, sexual maturation, mating behavior or reproductive capacity are found, clinical studies should not be undertaken in children unless additional appropriate studies show that ways of controlling such effects exist.

Since psychoactive drugs have a major site of action within the CNS, an extensive search is indicated for alterations of brain physiology and chemistry, neural development and function, learning, cognitive ability and behavior. Initial testing for drug effects on brain and neural functions should normally be completed concurrently with other animal testing prior to initial clinical studies. These CNS studies should include assessment in offspring of treated mothers as well as in the treated animals themselves. Specifically such tests should include gross examinations of whole brain and major brain regions for anomalies in development, measurement of brain weight, preparation and study of histologic sections through all major brain regions at each stage from fetal life to maturity to detect anomalies in patterns of cellular development and migration and the development of myelination. Brain DNA content, RNA content and protein content should be measured at the same developmental stages.

In addition there should be tests of behavior and learning on suitable animal models of human emotional behavior and learning. Since these evaluations are less well standardized, and often require extensive time periods, in most cases this part of the animal research can be carried out concurrently with initial clinical investigations, provided that the standard toxicity studies, studies of growth and development, and CNS studies have been completed. Acute and long-term effects of chronic drug administration on neurotransmitter levels and their metabolism should usually be included, since most psychoactive drugs appear to have significant effects upon the function of one or more neurotransmitters. Animal studies of similar duration should be submitted prior to the approval of human studies of that duration and should continue for the periods of time over which the drug is expected to be used in clinical practice.

CLINICAL STUDIES

Clinical drug evaluation typically progresses through four phases. Pediatric drug evaluations should include a similar progression of studies, although the categorization as to phase may differ somewhat from accepted adult clinical pharmacology. The stages of pediatric clinical drug evaluation should be as follows:

Initially (IA in outline below), short (2 to 3 days) single and multiple dose safety studies should establish initial dosage ranges which produce evidence of pharmacologic activity, side effects, or toxicity, and preliminary evidence of efficacy. Pharmacokinetic investigations are highly desirable at this stage to define blood levels elimination half-lives, and urinary excretion patterns of the drug and major metabolites.

Once short term safety information and some evidence of efficacy is available on pediatric patients, early (IB in outline below) pilot efficacy studies may be initiated jointly with longer duration safety studies and continued study of pharmacokinetics. These studies should provide information on potential therapeutic benefit utilizing appropriate dosage schedules derived from the initial single and multiple dose range and pharmacokinetic data. Typically these studies will be open (i.e., not double-blind) with placebo and/or standard drug comparisons as appropriate.

Utilizing information regarding safety and efficacy obtained in pilot studies, (II in outline below) doubleblind placebo-controlled studies should be performed on small groups of homogeneous subjects to establish objective evidence of efficacy. Safety studies continue but, depending on previous results, may be less comprehensive.

The final stage, (III in outline below) prior to marketing should include more extensive testing in controlled clinical trials involving more groups of patients in a variety of clinical settings selected to reflect the ultimate use of the drug. Where marketed drugs are available, comparison of the new agent to the existing drugs is essential at this stage. Long-term follow-up studies should also be initiated at this stage.

After release of the drug for general use, monitoring of drugs for adverse reactions and other special problems should continue along with long term follow-up studies to identify problems that may become apparent only after years of widespread use.

I. Initial Studies

IA. Early Studies

IA1. Initial Safety Studies

For many drugs, data may already be available from adult studies. However, when lacking, as for example for a drug intended only for use in children, initial single dose and short-term multiple dose studies should be conducted in adults to define a dosage range for human pharmacologic as well as toxicologic effects. Such studies in adults are usually conducted in normal subjects, although in some cases patient volunteers may be used.

The initial pediatric studies are designed to extend the adult pharmacologic and toxicologic data to children. Hence these studies should be conducted with small numbers (roughly 6 to 10) of pediatric patients, usually in pediatric clinical research units with facilities and personnel necessary for careful monitoring and for carrying out the pharmacokinetic, bioavailability and clinical studies. Initial safety studies are primarily to define initial dosage ranges which produce pharmacologic and toxic effects for use in designing subsequent efficacy studies. In children these studies should usually involve single dose or short-term multiple doses with an escalating dose schedule. Entry of patients into the study is often staggered to take maximum advantage of dosage information from preceding patients. Drug administration may vary from a single dose to a maximum of 3 to 5 days. Efforts should be made to obtain data on distribution and elimination half-lives, volume of distribution and bioavailability. Studies which establish evidence of pharmacologic activity and dosage ranges for pharmacologic and toxic effects should be completed before pilot studies assessing efficacy are initiated. Because the drug, at this stage of testing has not been shown safe for children, ethical considerations require that children in these initial safety studies be patients who might derive therapeutic benefit from the drug under investigation, either at the time of the initial safety study, or at a later date when more extensive studies will have established its efficacy. Evidence for behavioral effect should be sought during these initial safety studies but assessment of efficacy may be limited to carefully recorded clinical impressions by experienced investigators. Some consideration should be given to alternative designs which may be appropriate at this stage (Gehan, 1974).

IA2. Pharmacokinetic Studies

Pharmacokinetics may differ significantly in children, as compared to older persons. Furthermore, there may be important differences within the pediatric

population according to chronological and developmental age. Where techniques exist, rates of absorption, metabolism, active transport, distribution, and excretion must be evaluated in pediatric subgroups of various ages. Age-related effects may not bear a constant or predictable relationship to such standard clinical measures as height, weight, or surface area, etc., but rather to some other biologic marker of growth and maturation. Where possible, relationships between pharmacokinetics and relevant biologic markers should be assessed.

In addition, pharmacokinetics should be investigated in children with the particular disorder to be treated by the psychotropic agent under consideration. Abnormal metabolic conditions associated with specific syndromes and developmental disorders may uniquely affect the pharmacodynamics of a drug. For example, it has been suggested that infantile autism is associated with malabsorption. Thus, absorption, susceptability to toxic or adverse reactions, and rate of drug detoxification should be determined for the specific pediatric disorders in which the drug will be used whenever techniques have been developed.

The FDA General Consideration for the Clinical Evaluation of Drugs in Infants and Children (Appendix II) may be useful in selecting appropriate variables to be monitored and studies to be performed in establishing relationships between dosage, bioavailability, plasma drug levels and therapeutic and/or toxic effects.

IB. Pilot Studies of Efficacy and Safety

IB1. Investigator Selection

Initial studies of a drug in pediatric groups should usually be conducted largely in or in cooperation with pediatric clinical research units. Investigations should be carried out by experienced multi-disciplinary teams comprised of clinicians, psychopharmacologists and behavioral scientists. Clinical responsibility for the child's care should remain with a clinician(s) whose background and experience in treating the clinical syndrome and in child development provide assurance that adequate precautions are taken to insure patient safety. The sponsor should be prepared to provide information on the investigator's degree and area of responsibility for the patient's care, his experience with the clinical syndrome, his knowledge of child development, and his experience and training in the use of assessment devices.

IB2. Setting

The setting in which pilot-efficacy investigations are conducted should be selected on the basis-of availability of appropriate patients, ability to assure patient safety, potential for carrying out carefully done and well-documented studies by a multi-disciplinary investigative team. At this stage inpatient studies will usually be required. If the drugs are intended for eventual outpatient use, it may be desirable to continue observations regarding efficacy in the child's home environment once adequate assurance of safety has been established.

When initiation of drug treatment coincides with hospitalization or other alteration in the school, home and/or social environment, changes in behavior cannot be confidently ascribed to drug treatment. Hence the study must be designed to distinguish drug effects from effects of environmental change. This is usually best done by including a double-blind placebo control condition. When this design is not possible in the early stages, it should be

initiated as soon as possible. In any case, the degree to which the treatment setting differs from the child's usual home and school setting should be described.

IB3. Patient Selection Criteria

IB3a. Definition of Samples

Diagnostic classification of psychiatric disorders in children is less well-developed than for adult disorders. There is neither consensus nor established criteria regarding the nosology of childhood disorders. Examples of discussions regarding problems in psychiatric classification of children are included in Appendix (III).

Comparability and continuity between psychological disorders in children and adults cannot be assumed. Some disorders in childhood are continued into and diagnosed in adulthood. However, the classification of children's disorders must also provide for conditions which do not have an adult equivalent. Infantile autism and specific developmental disorders are examples. Further, a coincidence of labels between adult and childhood disorders does not necessarily mean either that the disorders are identical, or that they respond to the same pharmacological agents. Consequently, findings from adult psychopharmacological studies should not be extended automatically to the treatment of children.

It should also be noted that the same term may connote different disorders in children of different ages, and the same disorders may have different symptoms at different ages. Behaviors normal at one age may be a psychiatric symptom at another. These factors, along with the rapidity and variability in rate of growth and maturation complicate clinical evaluation of new psychopharmacological agents in children.

Children, more often than adults enter treatment involuntarily. It has been said that many children receive psychiatric treatment because they are "disturbing" someone rather than because they are "disturbed". Because there are ethical considerations regarding when children should be treated and when the intervention should be directed at their environments, care should be exercised in identifying appropriate clinical groups for psychotropic treatment. Input from responsible individuals from more than one environmental setting (e.g., home and school) is desirable to enhance the accuracy of the diagnosis of the child's condition.

Although psychodiagnosis of children is less well developed than we would wish, the diagnosis of the patient population from which subjects for study are selected should be specified because the symptoms which are the object of drug treatment may differ in cause, significance and drug response in the context of different syndromes. For instance, anxiety may present as a primary symptom or as a concomitant of depression. Similarly, hyperactivity may occur with equal severity as part of the hyperkinetic syndrome or as a symptom of psychosis. Diagnosis may utilize standard clinical nosology or a description of clinical phenomonology. In any case, diagnostic criteria must be defined fully enough to allow replication and should be consistent throughout a study. In addition to diagnosis, any criteria for entry into and for exclusion from the study must be described in terms that allow for objective measurement.

IB3b. Selection of Patient

Effectiveness of psychopharmacological agents can best be determined if the symptoms the drug is intended to treat are well specified. Instruments used to select patients should be sensitive to change in the target symptoms when treatment is effective.

Patient selection criteria should include specification of the context in which target symptoms occur. Cases should be selected to insure maximum diagnostic homogeneity uncomplicated by other psychiatric and/or neurological problems. Since the drug of interest may not be expected to affect all symptoms associated with a particular disorder, there may be symptoms which may occur in some eligible patients and not in others. The nature of the disorder and the knowledge of its natural history will determine whether the presence or absence of such associated symptoms must be controlled.

In addition to clinical syndrome or psychiatric diagnosis, a general set of variables which may affect treatment outcome in children is as follows:

1. Age of onset
2. Age of entry into study
3. Severity and description of symptoms
4. Duration and stability of symptoms
5. Associated symptoms
6. Sex
7. Socio-cultural-environmental context
8. Study context
9. Intellectual level
10. Prior treatment and response
11. Idiosyncratic response

Each patient's status should be completely documented with respect to each of these variables in pilot efficacy studies and samples should be as homogeneous as possible with respect to variables likely to affect the response of target symptoms to treatment.

These variables may be included in selection criteria of the sample or may be used as dimensions for stratification in the analysis of later efficacy studies. More detailed discussions of diagnostic and selection criteria can be found in Appendix III.

IB4. Exclusions

Criteria for inclusion and exclusion of patients in the study sample should be clearly stated before the study begins. These criteria should include statements regarding diseases, conditions, and other treatments (pharmacological and nonpharmacological) which, if present, would make

potential subjects unsuitable participants in the study. Issues to be considered in making these decisions are discussed in more detail in Appendix I.

Criteria should also be developed for dropping cases after they have been accepted into the study. These criteria may include placement in institution or foster home, family moves, onset of new illness, parents' decision to change physicians, emergence of side effects, child or parent's decision to discontinue, refusal to take medication as directed, refusal to cooperate with assessments, or beginning other medications or therapies that would have been grounds for initial exclusion. Every case that is dropped should be reported and documented. Consideration should be given to endpoint analysis of these data in final assessment of the study.

IB5. Other treatments

There are five common forms of treatment available for dealing with psychological and behavioral problems in childhood:

1. Special education (remedial reading, speech therapy, occupational therapy, special class assignment, resource room, etc.)

2. Counseling and psychotherapy (recipients may include child, family, teacher, etc.)

3. Environmental manipulation (institutionalization or changing family members, schools, classrooms, teachers, etc.)

4. Contingency management therapy (token economy systems, behavioral contracting, etc.)

5. Medication

In addition, pharmacological treatment of non psychiatric conditions may occur.

It is generally advisable to avoid all concurrent pharmacological treatments during this stage of study. Where possible, changes in other treatments should also be avoided. The greatest problem may be expected to arise when drug treatment coincides with a change in one of the other treatments. As mentioned previously, this cannot be avoided if the study required inpatient observations of children not already hospitalized. When a change in setting or treatment does occur, a sufficient baseline period of observation in the new setting or treatment, preferably with placebo, should be included before drug treatment is initiated. If the patient's symptoms decrease to, and remain at, a level below criteria for entrance into the study during such a baseline period, he should not be continued in the study.

Drug-free period. When safe and feasible, patients who have been on other drugs should have a drug-free period prior to starting the study medication. The length of this drug-free period will depend on the type and duration of prior medication and should be sufficient to remove all drug effect and outlast any withdrawal phenomena.

IB6. Variables to be evaluated.

IB6a. Efficacy

IB6a1. Variable Selection

Variables to be monitored in studying efficacy should be selected to (1) establish that the drug is being taken in scheduled amounts, (2) establish dosages necessary to produce behavioral effects and (3) determine that target behaviors are affected by the drug and not by some other intervention. To accomplish these goals, it is necessary to monitor not only target symptoms but also nontarget and nonspecific independent variables.

Target symptoms must be clearly stated and criteria for measuring their presence specified. These symptoms will ordinarily be those used to select patients into the study. Because measures of symptomatology differ in their sensitivity to drug effects, it is desirable to include several types of convergent measures of the symptoms selected.

Nontarget characteristics to be monitored should include situational and experiential/treatment variables discussed in Section IB5. Where possible, marker variables or measures known to be sensitive to drug effects should be included to monitor adequacy of drug intake.

For assessing effects on both target and nontarget behavior at this stage, it is particularly desirable to use measures which are relatively unaffected by repetition at frequent intervals. Suitable laboratory measures and observations are discussed in Appendices IV and V. Such measures can be used flexibly at frequent intervals during the initial stages of study.

IB6a2. Measurements

IB6a2a. Criteria for the selection of measurements.

The American Psychological Association Publica-tion "Standards for Educational and Psychological Tests" should be consulted for guidelines to selection, use and interpretation of psychological and behavioral measures. Procedures selected for use in drug studies should, whenever possible, meet standards of reliability and validity labeled "essential" in this publication. This includes laboratory and other procedures constructed specifically to study symptoms or behaviors under investigation.

In selecting instruments, reliability, validity, pertinence and sensitivity should all be considered. Reliability refers to measurement error and stability in test scores. Relatively unreliable measures may yield valid information when comparing groups of subjects yet be unsuitable for making valid inferences in individual cases. If the instrument selected for evaluating drug effects has units that are too gross for the degree and type of change expected, no drug effects will be detected even if they occur. In this instance, scores could be highly reliable and stable but the instrument would be insensitive.

Generally measures selected for use in defining characteristics of the study sample should have a high degree of reliability. Those used to measure change need to be sensitive to fluctuations in the range expected. For example, IQ tests are among the most reliable and stable measures available in the behavioral field. (Wolfensburger, 1961) They are, however, seldom sensitive to drug effects in the dosage range appropriate for treating children. Such tests may be very useful in defining the sample as mentally retarded or average intelligence but they will usually not be particularly useful as a measure of drug effects. Only instruments with acceptable levels of either interobserver and/or test-retest reliability (or both if appropriate) should be used in criteria for selecting patients into the study or defining pathological conditions.

Instruments used for measuring drug effects pose a different problem. Frequently it will be necessary to develop new instruments for measuring drug effects. However, whether reliability has been established or not, study design should be such that measures of change are not confounded with unreliability of drift in the measuring instrument. This is most effectively accomplished through use of a study design that employs randomized assignment of cases to active drug or placebo. However, placebo groups may be impractical in pilot studies. Whatever design investigators employ, they should be prepared to demonstrate that they have chosen and utilized behavioral measures in such a way that drug effects are not confounded with unreliability of measurement.

Validity generally refers to how well a test measures what it purports to measure or how well it predicts another variable. When a measurement is operationally defined in terms of the measuring device, there is little conceptual difficulty with validity. If anemia is defined in terms of the volume of packed red cells, one can point to the operation of packing the red cells to demonstrate that the hemotocrit is a measure of this volume. If, however, one is proposing to measure "cognitive functioning" or "vigilance", operational definitions are often much more narrow than is the conceptual use of the term. Interpretation of test results would be restricted to the level at which validity has been established. The Matching Familiar Figures, for example, is a widely used measure of "impulsivity" in children. It is not clear, however, whether this test measures impulsivity in the broad clinical sense or only in a situation characterized by response uncertainty and only when materials are restricted to pictures.

Five types of validity are traditionally distin-guished, face validity, criterion-related validities (concurrent and predictive), content validity, and construct validity (See APA Guidelines). Only the latter four of these are acceptable for interpretative inferences from test scores. Face validity, which is the mere appearance or proclamation of validity, even when based on a consensus of opinion, is an insufficient basis for interpreting the meaning of test scores. Both established measures and ad hoc

measures should be scrutinized carefully to assure that validity is not dependent upon face value alone. The American Psychological Association standards mentioned previously should be followed in determining whether measures meet requirements for content validity as opposed to face validity.

At a minimum, measures used in drug studies should (1) measure some dimension directly (as with height or weight or behavioral observations), (2) have appropriately developed content-referenced validity (as with symptom rating scales), (3) have criterion-related validity (as with laboratory measures of learning), or (4) some combination of these.

Considerations of reliability, sensitivity and validity will all contribute to determining how pertinent the measures are to the study. In addition, measures should be appropriate to the ages of the patients, to the symptoms under study, to the type of drug, to the severity of the patient's pathology, to the patient's IQ and social backgrounds, and to the conditions under which the drug will be used. Appendix I should be consulted for further details regarding factors to be considered in determining whether the selected measures are appropriate for measuring change.

When determining the suitability of measurements for the group under study, it is desirable to have normative data on samples of similar demographic characteristics and intelligence. Psychological and behavioral tests in particular often need different norms for different ages and socioeconomic levels.

Where certain scores on measures are to be used as criteria in selecting study patients, the necessity for normative data is particularly important. Age and sex are the most important demographic variables. Socio-ethnic background may also be of interest. The study should report how measurements of initial symptoms varied with these demographic variables. When there are significant correlations between symptom measures and demographic variables such as age and socio-economic status, it may be necessary to report results separately for these demographic categories.

IB6a2b. Types of measures

Typical procedures used in drug studies to evaluate efficacy have included: global ratings, symptom specific ratings, behavioral observations in natural or controlled settings, self-rating scales and investigator initiated measures. Assessment of both target and non-target parameters will ordinarily involve combinations of several of these types of measures.

Global ratings require judge(s), e.g., physicians, parents, teachers, peers, mental health workers, to rate the patient's status on one or more dimensions in one or more situations (school, home, clinic, playground). Agreement among raters

who observe the child in different situations is desirable but may not always be obtainable because patients are not consistent across situations. The amount of agreement among raters should be established, causes for disagreement ascertained, and where disagreements do not reflect actual differences in behavior, they should be reduced to a minimum. Ad hoc global rating scales can be useful but require validation. Specific scales are discussed in Appendix VI.

It is generally agreed that the reliability of rating scales is higher when the scale describes specific symptoms and behavior patterns rather than overall judgments. Scales may include ratings concerning the severity of behavior as in Conner's Teacher Questionnaire, the presence or absence of the behavior with differential weighting as in the Walker checklist or total number of behaviors displayed in each category as in Miller's scales. Rating scales applicable to school-age children are described in Appendix VI. It is anticipated that there will be only a limited need for studies of psychoactive drugs in preschool children; however, since there may be a need for such studies, Appendix VII has been included to describe rating scales available for use in the preschool period.

Self-report assessments have seldom been included in studies of the pediatric population. However, it is increasingly recognized that where feasible they can provide valuable information particularly in children with mental age above eight years. Self-report measures available for use in children are discussed in Appendix VIII.

Observations of pertinent behaviors by indepen dent observers who are naive regarding drug treatment studies and trained to an acceptable level of interobserver agreement may be the most desirable type of data for evaluating treatment effects, at this stage, particularly when drug trials are conducted in an inpatient setting. A number of scales developed for use in such situations are described in Appendix IV.

Measures of performance initiated by the investigator allow the creation of relatively standard situations for evaluating the child's functioning. Commonly these have involved three types: (1) psychometric tests, (2) laboratory tests and (3) clinical examination of social and emotional functioning in a psychiatric interview. Tests of intelligence and achievement are commonly used psychometric tests. Appendix V describes psychometric tests commonly used in drug studies. Reference texts described later in this section can be helpful in locating appropriate test forms for the patient group under study.

Experimental or laboratory procedures frequently lack adequate reliability data. Their clinical validity, if any, is sometimes difficult to establish. They may be used, however, if they are theoretically related to the type of behavior of interest and can be repeated frequently without altering validity of the results. They may be particularly

valuable in the absence of psychometric measures of the behavior under study. Ultimately such procedures and related theories may also contribute to establishing construct validity for important concepts in determining how drugs affect behavior. For example, the behavior of children children labeled hyperkinetic can be described in several ways. To determine which features are specifically susceptible to treatment with drugs, these features must be isolated and studied separately. A continuous performance test which requires vigilance (along with intelligence, recognition memory, general cognitive functioning, etc.) but which is not dependent upon activity level, may be useful in determining whether performance is disrupted by excessive activity per se and whether drug therapy alters performance with or without alteration in activity level.

Similarly, if it is postulated that drug therapy has an effect on "learning", the investigator may be interested in using laboratory procedures such as paired-associated and serial learning to evaluate this effect rather than an intelligence test based on accumulated past learning over long periods of time. Unfortunately, without empirical support, results of these procedures cannot be assumed to correlate with learning in real-life situations (predictive validity) nor even to represent the same domain as learning in school (content referenced validity). Measures of different aspects of the same procedure (trials to criterion, total errors, latency of responding) may not even correlate with each other, much less with other learning procedures.

Hence, results of these procedures must be interpreted cautiously with respect to their meaning for individual patients. Protocol design must be appropriate to account for errors of interpretation that may result from use of procedures which have not been subjected to the rigorous scrutiny involved in the development and use of published tests. Continued development and use of such procedures is to be encouraged since they may ultimately contribute to an understanding of diagnostic patterns and/or to be helpful monitoring response to medication. In Appendix V common psychometric and laboratory procedures previously used in pediatric drug research are discussed.

Social and emotional functioning is most commonly assessed through ratings of behavior in clinical interviews, playroom observations and test situations along with information from projec-tive tests or other personality tests. These are discussed more fully in Appendix IX.

Environmental assessment can be limited to reports of demographic variables such as family composition, parents' income, education, occupation, and of school placement or expanded to include ratings of environmental support systems and assessment of family attitudes and characteristics. Similarly, school influences can be assessed through descriptive information on program assignment, classroom type (open, unstructured) and size, methods of instruction or assessment of teacher variables. Appendix X

presents a review of procedures for use in environmental assessment.

Often several types of instruments may be available for each group of psychological, behavioral, and social parameters. These vary in source of information, measuring units, and objectives as well as reliability, validity, and sensitivity. Usually, more than one type of measurement for each parameter is desirable. However, all available techniques need not be employed in each study.

Measurement of change will involve a comparison of measures over time for example baseline in comparison to post-treatment. This is often accomplished through comparison of baseline scores with post-treatment scores on tests that have been prospectively selected and recorded throughout the study. Measuring devices will differ in the readiness with which they lend themselves to various types of change analysis and vulnerability to confounding with other variables such as practice effect. Devices for measuring change should be chosen to be sensitive to change within the framework of the experimental design being planned. This is discussed more fully in Appendix I.

A multitude of instruments are available for assessing a variety of behavioral and psychological parameters. Only a small fraction of those available have actually been utilized in drug studies in the past. In addition to those discussed in the attached Appendices, reference texts which describe a number of procedures in detail along with descriptive and evaluative reviews are listed below:

Buros, O.K. Mental Measurements Year book. Updated periodically, the most comprehensive source of its kind.

Bommarito, O.G., & Johnson, J.W. Tests and Measurements in Child Development; a handbook, Vol I and Vol. II. Jossey-Bass, San Francisco, 1971. Includes many procedures not included in Buros along with references. Primarilydescriptivereviews. Limitedevaluationinformation.

Frankenburg, W.K., & Camp, B.W. Pediatric Screening Tests. C.C. Thomas, Inc., 1975. Descriptive and evaluative reviews of screening procedures available for use in detecting abnormality in a variety of physical, sensory and psychological areas.

Walker, D.K. Social and emotional measures in preschool and kindergarten children. Jossey Bass, San Francisco, 1973.

Comrey, Becker, & Glaser. A Sourcebook for Mental Health Measures. Human Interaction Researc Institute, 10889 Wilshire Boulevard, Los Angeles, California 90024.

A special issue of the Psychopharmacology Bulletin (1973), entitled "Pharmacotherapy of Children" (DHEW Publication No. (HSM) 739002) contains recommendations regarding a

standard battery for use with children in the Early Clinical Drug Evaluation Units (ECDEU) Program. Portions of this publication have been reprinted as appendices to the present document.

These references can provide information for locating procedures which may be useful in drug studies but which have not been tried previously.

IB6b. Safety

IB6b1. Variable Selection

Establishing safety usually involves monitoring nontarget characteristics or aspects of functioning with which drug therapy may be expected to interfere. There are three broad categories of such parameters: behavioral and psychological status, physical growth and development, general physical physiological status.

Behavioral and psychological status should be monitored through evaluation of general intellectual, cognitive, social and emotional functioning whether or not these include the target symptoms. This evaluation should include assessment of learning in both laboratory and real life settings. Considerations discussed under Efficacy regarding criteria for selecting measures and types of measures also apply here.

Physical growth and development including sexual maturation should be monitored routinely. Rate, amplitude and timing of physical growth and development vary among normal children. Hence evidence that a drug causes alteration in growth may be difficult to establish. However, it is unusual to find major deviations from family patterns in groups of normal subjects. At a minimum, the child's height, weight, head circumference and segmental proportions (span; upper: lower segment ratios) should be recorded at intervals of 3 to 6 months. These should be plotted on suitable growth graphs (Frankenburg & Camp, Pediatric Screening Tests for reviews of available norms), and if there are abnormal findings, comparisons with the growth of other family members should be made where possible before assuming that the drug is either responsible or not involved.

Ages at which various indices of sexual maturation (pubarche, adrenarche, menarche) occur should be recorded and compared with ages of their occurence in other family members. The Tanner scale (Tanner, J.M., Growth and endocrinology of the adolescent. In Gardner, L.I., Endocrine and genetic diseases of childhood. W.B. Saunders, Philadelphia, 1969, pp. 19-60) is an example of a method which can be used for assessing stages of puberty and the progression of these stages. As with growth, normal subjects seldom show major deviation from the family pattern of progress through the stages of puberty.

Assessment of skeletal maturation through serial bone age films should be included when drugs are chronically administered or if animal studies suggest effects on growth or skeletal maturation.

Assessment of general physical and physiological status should include (1) a standard medical history, (2) physical (including neurological) examination for side effects, and (3) special procedures to monitor hematologic, hepatic, renal, cardiovascular and endocrine status. Tests of central and peripheral nervous system function and other aspects of physiological and metabolic status should be included when indicated. These additional parameters can be selected on the basis of clues taken from studies in adults and from knowledge of the pharmacologic and chemical nature of the drug.

The standard history and physical examination form used should be followed completely for every subject. In Appendix XI a suggested form for eliciting and recording side effects is presented. Either this or a similar procedure should be carried out. The procedures used should be fully reported.

Possible screening procedures to be included in monitoring the various organ systems are described below. Those on which baseline measures should be done routinely are starred (**).

Cardiovascular - **Blood presssure, **heart rate, **electrocardiogram, blood cholesterol and triglycerides;

Hematological - **complete blood count including differential and platelets, **G-6PD deficiency screening;

Renal and metabolic - **routine urinalysis for specific gravity, protein, glucose, ketones and microscopic examination; **urinary amino acids; renal clearance (**creatinine, PAH, inulin). **2 hour p.c. blood glucose, **blood urea nitrogen, **blood pH and electrolytes including calcium and phosphorus.

Hapatic - **Bilirubin, **SGOT, **LDH, **alkaline phosphatase, **total protein and serum electrophoresis, BSP.

Endocrine - **bone age, thyroid function tests (TSH, T3R, T4) growth hormone, Serum LH and/or FSH, testosterone cortisol; cytology for estrogen effects (females); urinary 17-Ketosteroids, VMA.

Central nervous system - Electroencephalogram with cortical evoked potentials and/or frequency spectrum analysis.

Peripheral nervous system - Electromyography and nerve conduction, CPK.

It is extremely important that growth measurements, particularly height and weight, be measured in a standard manner with the child stripped. Although small deviations from expected growth over short periods (3 to 6 months) are of little clinical significance in evaluating drug effects on growth, they may be used as a basis for determining how endocrine status should be followed. Even small deviations from expected growth over 3 to 6 months may be cause for evaluating thyroid function, growth hormone and bone age. Similarly, delay in sexual maturation or suspected drug influence on gonadal function should lead to monitoring of serum LH and/or FSH, testosterone and estrogen and, in the case of females, cytology for estrogen effect. Evidence of premature sexual maturation would prompt, in addition, measurement of urinary 17-Kestosteroids.

ACTH and cortisol levels may be indicated when alteration in blood electrolytes occur.

Screening tests for specific endocrine function may produce positive answers which then have to be pursued, but which have no clinical significance. For instance, numerous drugs are known to alter measurements of thyroid function without evidence that thyroid function is actually altered. Unless previous studies on animals or adults provide a basis for suspecting some endocrine effect, monitoring growth may be the only practical way to decide whether endocrine studies are worth pursuing.

The importance of monitoring various aspects of central and peripheral nervous system functioning will also depend on the type of drug, its expected effects and information about the value of these procedures in similar situations. For example, the cortical evoked potential is currently useful in evaluating neural functioning against norms available for healthy children, those addicted to drugs and those with hearing and visual impairment.

IB7. Schedule and frequency of assessment

The scheduling of assessments should be relevant to the potential target and nontarget effects of the drug. Points to be considered in determining this schedule include the speed with which the drug is expected to produce changes, the duration of effects, the stability of the variables under investigation, the level of remission expected, the nature of the measuring device, and the type of data analysis planned. For example, if one employs a single-subject cross-over design in which the criterion is changes in behavior in an activity room, observations might be scheduled daily over a period of 2 weeks. If the investigator wishes to utilize a time series analysis as proposed by Glass, Wilson and Gottman (1975) to study the time course of drug effects on behavior in a classroom, continuous observations over several hours might be needed each day to obtain sufficient data points for this type of analysis.

In planning for monitoring safety, studies of the drug in adults should be searched for suggestions as to important parameters to be assessed. In addition, baseline measures should be obtained on those physical and laboratory procedures identified by (**) in Section IB6b1 for routine analysis. Except for G-6PD, measures of drug effects on these functions should be repeated at 24 and 48 hours after initial administration of the drug in children and followed at appropriate intervals throughout the study. Vital signs should be monitored frequently within the first 24 hours.

IB8. Duration of trials

Duration of a study will depend on the disorder being treated and the pharmacological properties of the drug. Whenever appropriate pilot trials should be carried out over several weeks.

IB9. Study Design and Control Procedure

Pilot studies in children should generally consist of open studies of small groups (10 to 15) of patients followed intensively while relationships between dosages, bioavailability, blood levels, therapeutic and/or toxicologic effects are being established. These patients may be the ones who participated in initial studies of safety and pharmacokinetics if they meet selection criteria. These early studies should be designed to allow investigators sufficient

flexibility to explore all relevant aspects of a drug's activity and to establish dosage range for later use in double-blind studies.

Often these studies may employ single subject designs. Even at this stage, however, placebo controls are very desirable, either in the form of a cross-over design, or a time series analysis (Campbell & Stanley, 1966).

By appropriate training and blinding of data gathering personnel, it is sometimes possible, even at this stage, to obtain unbiased measures of behavioral or psychological change which can be analyzed in the manner of later double-blind studies. Where evaluation is by subjective judgment in an open study, it will not be possible to rule out observer bias. However, hypotheses generated by such studies may be subjected to proper experimental test in later, double-blind studies. Later comments regarding types of data, analysis of change and statistical vs. clinical significance are also applicable here.

While the design should allow flexible dosages at this stage, different investigators should all use the same measures and patient selection criteria.

II. Shortest Duration Studies to Establish Clear-cut Evidence of Therapeutic Potential and Safety.

II.A. Investigator Selection

Selection criteria for investigators during this stage should be similar to criteria during pilot- efficacy studies. Studies should be carried out by a multi-disciplinary investigative team which includes some members not otherwise involved in clinical care of the study patients. This procedure permits some data gathering by individuals who are maximally objective.

II.B. Setting

Studies should be carried out in cooperation with specialized pediatric clinical research units with facilities and personnel to perform appropriate safety and clinical monitoring. Each study should clearly describe the treatment and setting and the subject's living and school situations. Studies at this stage should be carried out in at least three different centers. (See Appendix I for a discussion of how to proceed when results obtained from these three centers are not congruent.)

Where possible, the settings of Phase II studies should be representative of the settings in which the drug is expected to be used. If early Phase II studies require the child to be removed from his/ her natural environment for initiating and monitoring drug therapy, studies should be conducted as soon as possible with the child living and going to school in the environment in which drug therapy is expected to be used. Such studies in the natural setting should be completed prior to initiating Phase III studies.

II.C. Patients Selection Criteria

Comments under IB3 are applicable here.

II.D. Exclusions

Comments under IB4 are applicable here.

II.E. Other treatments

Comments regarding definition of "other treatments" in Section IB5 are applicable here. The child's status with respect to these other treatments should be monitored carefully throughout the study. Concurrent chronic pharmacological treatments (neuropsychiatric and nonpsychiatric) should generally be avoided at this stage.

While it may be neither realistic nor desirable to avoid all nonpharmacological treatments (e.g., special education), studies at this stage should be designed to distinguish drug effects from effects of "other treatments" including effects observed when drug therapy coincides with initiation or change in one of these other treatments or change in the family structure. Patients who have been on other drugs should have a drug-free period prior to starting the study medication. The length of this period will depend on the type and duration of the prior medication.

II.F. Parameters to be evaluated

II.F1. Efficacy

Assessment of change should involve not only the target symptoms or behavior for which the drug is being administered, but also the development of side effects if any, and the degree of change in non-target characteristics selected to monitor the child's general status. The classic problem in this regard is that a drug which effectively assists in decreasing undesirable behavior may also so sedate the patients that he/she is unable to function well. The characteristics to be monitored must be selected to answer the question of whether the drug is effective, whether rival, plausible hypotheses can explain effects which are observed, and whether ill effects of the drug occur.

By this stage in drug development, pilot studies should have given indication of possible therapeutic effect, important behavioral and pharmacological effects and should have eliminated drugs which have little or no psychoactive effect in children or which have unacceptably high toxic potential. In addition, Phase I studies can be used to pin-point the target variables which should be studied further. However, monitoring of nontarget variables should continue as previously.

Previous comments regarding selection and use of measuring devices are applicable here. The investigators should be prepared to demonstrate that they have adequately controlled for instrument unreliability either by choosing reliable instruments and using them in a standard manner or through study design.

II.F2. Safety

Monitoring of acute and chronic toxicity should depend upon results of Phase I studies and total available knowledge of the drug. Specific types of toxicity suggested during acute safety studies should be studied further in early pilot efficacy studies. Additionally, safety monitoring during Phase II would include all of the parameters cited in Section IB6b.

In order to collect information on possible long term drug effects on growth, development and sexual maturation, it would be desirable for patients enrolled in Phase II studies to agree to continue long term contact with the investigators. This will lay the groundwork for contacting children exposed to investigational drugs later in the event late onset effects are suspected whether or not the drug ultimately reaches approved clinical use.

II.G. Schedule and frequency of assessment.

Results of pilot studies will be helpful in deciding on types and frequency of measurement. The type of data which is being collected and the type of analysis planned will also help to determine the frequency of assessment. Presumably some measures adapted for frequent assessment such as several times daily may be replaced by measures which are performed less often.

For example, behavior observations performed by independent observers several times a day during pilot studies may be replaced by a rating scale completed several times a week by the child's care taker.

II.H. Duration of trials

Duration of the study will depend upon the disorder being treated and the pharmacologic properties of the drug. However, in most instances investigations should be carried out over a minimum of several weeks and, where possible, longer. It is desirable to continue studies for a minimum of 6 months when drugs are intended for chronic use over several months or years. This is particularly important in pediatric psychopharmacology because long term drug effects on growth, development, learning and maturation are equally important to establish as drug efficacy itself.

II.I. Study Design and Control Procedure

Studies should use samples which are homogeneous with respect to diagnosis, severity and other relevant variables, and should use double-blind techniques, control conditions and random assignment to treatment groups. For many pediatric psychoactive drugs the double-blind experiment with a placebo condition will be the preferred approach. Consideration should also be given to inclusion of an initial 2-week placebo "washout" period to eliminate placebo-responders and determine the effects of repeated measures.

When the study group includes severely ill children an active comparison drug rather than placebo control may be considered. Where a serious disorder with fairly stable symptomatology is being investigated, a crossover design may be employed as well.

Methods of assignment of patients to treatment groups may vary depending on the phases of the investigation and the contemplated sample size. When sample sizes are small, random assignment alone will not necessarily insure comparability between groups on critical variables such as age, sex, duration of symptoms, severity and other considerations discussed under Patient Selection Criteria. There are many techniques which allow random assignment to index and control groups and yet maintain comparability between groups (e.g., selecting matched pairs who are then randomly assigned, or setting up quotas, and taking the next patient who meets the quotas, etc.). Earlier studies should be investigated for suggestions as to which variables need to be controlled

At this stage, it is desirable to include studies of dose-related effects and contrasts between the new drug and a standard drug of known efficacy as well as contrasts between new drug and placebo. In each instance, protocol design must be such that plausible explanations of changes due to nondrug factors can be ruled out. Campbell and Stanely's (1966) Experimental and Quasi-experimental Designs in Research may be consulted for details regarding recognized sources of invalidity in conclusions which result from defects in protocol design.

The sample size needed to establish effectiveness will depend on purposes of the study, the expected magnitude of drug effects, the desired probability that effects will be detected and the type of data and data analyses planned. It is rarely possible to achieve reasonably definitive answers with less than 20 patients per study, and considerably more may be required. Detailed discussion of considerations regarding sample size are available in several reviews included in Appendix I.

Data may be both qualitative and quantitative. Qualitative (categorical) data usually consist of enumerations or counts of various categories of patients status or change. Quantitative measures are based on the assumption that the variable under study is distributed in infinitely varying amounts in different patients. Many psychological measures use frequency counts as the fundamental datum, e.g., number of errors, number of times an act occurs in a particular time period. While frequency counts are technically qualitative, they can often be treated as quantitative measures without serious error.

Appendix I presents a more detailed discussion of the differences between qualitative and quantitative data. Parametric and nonparametic statistical techniques should be used as appropriate and multivariate analytical procedures to summarize complex data.

Efficacy should be claimed only for the population represented by categories of patients who have shown significant therapeutic responses. It is important, therefore, to document the characteristics of the sample in sufficient detail to identify the subgroups treated effectively. In addition, it is important to know what specific aspects of a disorder are affected by the drug. This requires documentation which will permit analysis of which symptoms improved and which did not. Findings of efficacy should distinguish between drug effect and drug efficacy. Drug effects include efficacy - i.e., improvement in target symptoms - but also include changes in associated symptoms or marker behaviors. Global ratings may be misleading because they may reflect drug effects without significant improvement in target symptoms.

Statistically significant differences obtained in planned contrasts represent only the minimum evaluative statement which should be presented in support of drug efficacy. For results to have clinical significance they must reflect an improvement in target symptoms, and be of such magnitude as to reflect improved function in the "real world."

Investigators should provide evidence pertinent to establishing the clinical significance of their findings as well as statistical significance (see Appendix I).

In long-term uncontrolled studies of drug effect in children, one of the most critical problems in assessing clinical significance is determining whether drug treatment produces changes over and beyond those which can be expected on the basis of maturation alone. Cross sectional studies at

different ages do not provide the information regarding changes which may be expected within a child (Shale, 1965).

General problems and pitfalls of research designs proposed to measure effects of treatment superimposed on maturational processes are discussed in the previously mentioned work of Campbell and Stanley (1966). Additions to the original Campbell and Stanley Monograph are summarized by Wortman (1975). Kenny (1975) has recently discussed indications and contraindications for various ways of measuring change, and Downing, Rickels, Wittenborn and Mattson (1971) have discussed this problem extensively in relation to assessing the effectiveness of psychotropic agents. These references should be consulted for guidelines to evaluation of the appropriateness of data analysis to the study design and types of data collected.

III. Extension of Efficacy and Safety Studies

III.A. Investigators

On the basis of previous studies it should be possible to eliminate some of the safety measures. However, a multi-disciplinary investigative team may still be required.

III.B. Setting

These studies should be carried out in the child's usual setting.

III.C. Patients Selection Criteria

Previous comments regarding patients selection are applicable. However, these final, pre-marketing studies should reflect the spectrum of patients and symptom patterns encountered in the type of clinical practice in which the drug will be used. In these larger studies, it is still important to characterize the sample with sufficient detail so that stratification of patients can be carried out in analysis, and subgroups with differential drug or treatment effects can be detected. Data on patient characteristics should be uniformly elicited, rated and recorded during sample acquisition.

III.D. Exclusions

Previous comments regarding exclusions are applicable here with the addi tion that information obtained from previous studies may be used to establish different criteria for exclusion. As previously stated, specific criteria for inclusion and exclusion must be stated prior to initiating the study. Exclusion criteria should not be so restrictive that children often referred for treatment are excluded from study.

Drop-outs must be carefully recorded and the reason for dropping described in detail. Previous comments regarding criteria for dropping from the study and end-point analyses remain applicable.

III.E. Other treatments

At this stage, treatment with other psychoactive drugs should be avoided but treatment with nonpsychoactive drugs known to be free of behavioral effects may be permitted. Greater flexibility may also be permitted in the type of control exerted over other nonpharmacological treatments. It is assumed that the larger sample sizes and random assignment of cases to active drug and control groups will eliminate bias that might be introduced by these other

treatments. However, other nonpharmacological treatments should be carefully monitored.

III.F. Parameters to be evaluated and frequency of Assessment

III.F1. Efficacy

Previous comments are applicable.

III.F2. Safety

Acute toxicity monitoring during this stage may be less extensive than during earlier studies and will depend on the particular properties of the drug as determined in this earlier testing. Although less intensive safety monitoring may be indicated, it should include the routine hematological, hepatic, renal and cardiovascular measures listed in Section IB6b1.

III.G. Duration of Treatment

Many childhood psychological disorders are chronic. Since psychotropic agents may have to be administered over extended periods of time it is important to assess both long and short term effects. Short term trials extending over several weeks or months should be carried out as in previous studies. Some trials should last long enough so that evaluation of habituation potential, development of tolerance, and effects on the processes of development and maturation can be made over the anticipated extended period of treatment. Appendix I should be consulted for further discussion of this issue. Because some effects which may not occur during one phase of development may appear during others and adverse effects may occur in patients with certain disorders but not in others, generalizations from the populations under study will have to be probabalistic.

Because of the likelihood that an investigational drug which reaches this stage of study will undergo widespread clinical use and ultimately FDA approval, consideration should be given to assessment of adverse, late onset effects in an adequate and representative sample of patients participating in these final pre-marketing studies, particularly if the drug may be used chronically. The justification for and difficulties of monitoring late onset effects are discussed more fully in the General Guidelines published by the American Academy of Pediatrics which should be consulted for details. Appropriate methods of followup for these effects should be based on the nature and use of the drug, its pharmacologic effects and age of the patients at the time of drug exposure.

III.H. Study Design and Control Procedure

These studies will extend the investigations of drugs which have shown promise in earlier studies to larger, more standardized dosages and treatment settings. Such studies should consist of double-blind, placebo-controlled studies with random assignment to experimental and control conditions and strict adherence to principles of sound experimental design and protocol. Where an existing drug is available for comparison, some studies at this stage should include comparison of the new agent with the standard. Samples should be carefully characterized to establish all pertinent baseline measures and to permit comparison of treatment groups for absence of bias in the selection process. With few exceptions, each study should usually include a minimum of 30 to 40 patients.

Stratification of the study samples on some key variables may permit subgroups of patients to be identified that respond well or poorly to this psychoactive agent, and thus allow developing predictors of drug response.

Previous considerations regarding data analysis are applicable here.

REFERENCES

Campbell, D.T., & Stanley, J.C. Experimental and quasi-experience designs for research. Chicago, Illinois: Rand McNally & Company, 1963.

Downing, R.W., Rickels, K., Wittenborn, J.R., & Mattsson, N.B. Interpretation of data from investigations assessing the effectiveness of psychotropic agents. In J. Levine, B.C. Schiele, & L. Bouthilet (Eds.). Principles and problems in establishing the efficacy of psychotropic agents. (Public Health Service Publication No. 2138) Washington, D.C.: U.S. Government Printing Office, 1971.

Frankenburg, W.K., & Camp, B.W. Pediatric screening tests. Springfield, Illinois: Charles C. Thomas, 1975.

Gardner, L.I. Endocrine and genetic diseases of childhood. Philadelphia, Pennsylvania: W.B. Saunders, 1969.

Gehan, E.A., & Freireich, E.J. Non-randomized controls in cancer clinical trials. New England Journal of Medicine, 1974, 290, 198-203.

Glass, G.V., Willson, V.L., & Gottman, J.M. Design and analysis of time-series experiments. Boulder, Colorado: University of Colorado, Laboratory of Educational Research, 1972.

Kenny, D.A. A quadi-experimental approach to assessing treatment effects in a nonequivalent control group design. Psychological Bulletin 1975, 82, 345-362.

Shaie, K.W. A general model for the study of developmental problems. Psychological Bulletin 1965, 64, 92-107.

Tanner, J.M. Growth and endocrinology of the adolescent. In L.I. Gardner (Ed.) Endocrine and genetic diseases of childhood. Philadelphia: W.B. Saunders, 1969.

Wolfensberger, W., & Menolascino, F. Basic considerations in evaluating ability of drugs to stimulate cognitive development in retardates. American Journal of Mental Deficiency 1968, 73, 414-423.

Wortman, P.M. Evaluation research: A psychological perspective. American Psychologist, 1975, 30, 562-575.

INTRODUCTION TO APPENDICES

The following Appendices are intended to amplify procedures and measures that have been discussed more generally in the main body of the Guidelines. It will be seen that there is a diversity of approach which represents the diversity of the field. Opinions expressed are the author's own, and do not signify specific recommendations from the FDA and its consultants.

What is intended, however, is a sense that careful measurement in any one of these areas will be indicated for some studies, and that it is likely that a combination of both global clinical assessment as well as objective data will be needed for diagnostic description and documentation of clinical change.

APPENDICES

APPENDIX I

GENERAL PRINCIPLES FOR PSYCHOACTIVE DRUG STUDIES*

These guidelines are expected to be of assistance to those who must plan, conduct, interpret, and eventually review programs of study designed to establish the safety and efficacy of psychotropic substances.

Guidelines are only suggestions which will not be sufficient to the requirements of some situations and should never be regarded as absolute or obligatory criteria. The present guidelines are intended to reflect the current consensus of a responsible group of clinical scientists and are presented with the belief that the quality and pertinence of Phase III research would be generally improved if the guidelines were considered during the planning and conduct of large scale clinical inquiries.

Investigations are never complete or sufficient. They cannot answer all the questions that might legitimately be raised, and they may not be viewed with unreserved approval by all who examine them. Thus by their nature, inquiries are characterized by faults. It is useful, however, to distinguish between two kinds of faults.

The most obvious fault is bias, where the conditions of the inquiry favor or handicap one of the agents under comparison. If the biasing influence operates against the compound under test, the therapeutic effect is obscured by the biasing influence. If the investigational compound is found to be superior to placebo despite the burden of uncorrected detracting bias, the efficacy shown must be accepted at face value despite the fact that even greater superiority might have been shown had the conditions of study been less prejudicial. Biasing influences which systematically favor the agents under appraisal must be recognized and corrected in some satisfactory manner. If such favorable bias is uncorrected, the claim to efficacy may be denied.

A second fault of investigations is their insensitivity. Insensitivity accrues from many sources. The criteria may have been intrinsically unreliable or unreliably applied. The criteria may not have been fully expressive of the central therapeutic effect and, in this sense, lacking in validity. The conditions of treatment may have been somewhat inimical to the requirements of the investigation, and diminished precision may have accrued from many sources. Confounding influences, therapeutic and otherwise, may have served to obscure the contrast between the active principal and the placebo, and there may have been various major sources of heterogeneity which were either unrecognized or uncorrected in the statistical analysis, thereby reducing the sensitivity of the tests of significance. There is real hazard that the efficacy of a potentially valuable agent may be obscured by these various sources of insensitivity.

Inquiries conducted in a realistic clinical context inevitably involve faults of the kinds suggested here. Guidelines are offered on the assumption that some of the faults which bias or obscure clinical inquiries are preventable. If the sources of fault are anticipated and their presence recorded, their influence can often be reduced in the handling of the data. Clinical trial findings must be gauged in terms of evidence of efficacy and not in terms of detracting technical faults of the research.

*Written by J.R. Wittenborn, Ph.D., Rutgers University, New Brunswick, New Jersey.

A. DEFINITION OF SAMPLES

When a new compound is submitted to clinical trials, it is assumed to be efficacious in the management of certain indications as they occur in some clinically definable group of patients. As new drugs are being proposed for psychological and behavioral disorders questions arise concerning the definition of the condition for which efficacy is claimed. Often these conditions do not correspond with any familiar diagnostic stereotype. Hence, they must be defined comprehensively not only in terms of the manifestations for which remission is sought, but also in terms of the symptomatic, situational, and historical context in which the manifestations occur and are effectively treated.

Traditional diagnostic statements cannot always be expected to correspond with the indications presented for treatment. To claim efficacy throughout a diagnostic entity on the basis of a desirable response among some unidentified but limited portion of patients may be disadvantageous for both the drug house and the patient. Under such blanket claims many patients are treated ineffectively. Thus analyses that seek to identify the portion of the sample which responded best should be regarded as a responsible attempt to identify the portion of the sample appropriate for the treatment under consideration. If the optimally responding portion of the patients is bound to be similarly identified in several of the independent studies comprising a series, the mutually confirming findings are a proper guide to the appropriate use of the medication and the claim of efficacy for the responding portion of heterogeneous samples should not be interpreted as a post hoc use of adventitious factors to sustain a limited claim for efficacy. With this confirmation the medication can then be recommended for only those patients who can be expected to respond.

B. THE DOCUMENTATION OF THE SAMPLES

In planning clinical trials to establish the efficacy and the safety of a new drug, it is important to acknowledge that there will be differences between the sample sought, the sample obtained, and the sample effectively treated. The population for which therapeutic efficacy is eventually claimed should be limited to the population defined by the patients effectively treated. Thus, suitable documentation of the nature of the samples is essential.

The documentation of the sample will require a diversity of information, including pretreatment measures of criteria of therapeutic effect. Much of the documentating information obviously cannot be used as criteria, however. Proper documentation of the sample should include at least the following eight kinds of information.

1. Pretreatment Symptoms

Since investigators involved in a Phase III Investigation may not subscribe to or follow the same diagnostic criteria, a definition of the sample effectively treated requires a substantial and detailed body of standard information on every patient included in the trials.

It is useful, therefore, to distinguish between the target symptoms that are the object of treatment and the accompanying symptoms required for the clinical definition of the sample. The pretreatment use of a comprehensive symptom rating scale is recommended to define patient populations which do not correspond exactly with any generally accepted grouping. This standard information helps to document the degree of disturbance in areas other than the one(s) under treatment. It helps to specify whether the effective management of a group of target symptoms occurs in the presence or in the absence of other defineable problems, for example, mental retardation.

2. History of Prior Episodes

In episodic disorders, the history of prior episodes often bears a relationship to the response to medication. For example, the number of prior treatments, the age at which the first episode occurred, and the manner in which the patient responded to prior medication can all be relevant to the patient's response to current treatment. Such information is essential to proper documentation.

3. History of Current Episode

Whether the current symptoms occurred without obvious precipitating provocation can be an important distinction, as well as the duration of the current episode and the response to other medications in the course of the current episode. Diagnostic statements should be included as well. Whether the patient has had or is having psychotherapy for the current episode is also important. Such information is essential to proper documentation.

4. The Clinician-Investigator

The kind of setting in which the present study is conducted, e.g., public supported clinic, private clinic, private psychiatric practice, or nonpsychiatric practice of medicine, is known to be important in determining the response of patients to psychotropic medication. Important also is whether the investigator customarily develops a therapeutic relationship with his patients and whether this is a part of a specific psychotherapy, handled indirectly in group psychotherapy, or merely a quality of the investigator's concern and interaction with patients. The experience that the investigator has had with psychotropic medication and his confidence in medication is important. It is useful also to know whether the patients are supported by insurance which provides only a limited period of treatment, e.g., 28 days. An additional useful part of the documentation would include the investigator's attitude toward, experience with, and manner of use of the assessment devices.

5. Concurrent Treatments

Although all reasonable efforts should be made to avoid concurrent treatments of any kind, some appear to be inevitable. It is most important, therefore, to provide complete documentation for any concurrent therapeutic influence. It may be in the form of group psychotherapy or social work counseling, either related to or independent of the investigator's treatment. It is possible also that for some patients night time sedation may be available during the first few days of treatment. Not infrequently investigators will feel free to apply a concurrent psychotropic of a different type, e.g., major tranquilizers, stimulants or sedatives. The concurrent use of oral antihistamines for some other indication can also be of interest as can the use of thyrotropic hormones. Sometimes the patient is suffering from a significant organic disorder, and the behavioral symptoms may be related to the physical state or the treatment as with phenobarbital for seizure disorders. Occasionally this is discovered only from a scrutiny of the uses of concurrent medication. Although concurrent medications are usually assigned independently of whether the patient is on the experimental medication of the placebo control, it is important to maintain a complete log of all concurrent medication, including treatment for headaches and colds.

If the concurrent medication was distributed among patients receiving the investigational medication and patients receiving the control medication in a uniform manner, its probable effect would be to obscure the potential differences between the medications. A similar effect could be expected if the concurrent medication were selectively assigned to supplement the weaker treatment. Regardless of the fact that concurrent medication usually tends to diminish the contrasts between an active medication and an inactive control, cases receiving concurrent treatment and

claims based upon them may be disallowed by the FDA. For these and other reasons, cases receiving concurrent medication are left in the sample at the peril of the investigation.

6. Demographic Variables

Age has been found to play an important role in the relative efficacy of psychotropic medications, and other demographic factors of possible relevance include sex, race, social class status; all are a part of proper documentation.

7. Sample Size and Heterogeneity

The probability of showing a statistically significant effect is a function of the magnitude of the effect, the size of the sample, and the heterogeneity of the sample. These factors can be only partially controlled by the investigational plan, however. There are inevitable differences between the sample sought and the sample obtained. In clinical investigations many considerations intrinsic to the clinical situation and unrelated to the treatment per se can result in substantial attrition in the size of the sample actually obtained, and it is difficult to anticipate the magnitude of the effect, even when relatively well known compounds are being examined.

The investigational plan may provide for a restriction or for stratification of some sources of heterogeneity. In studies of mixed states of anxiety and depression, particularly, heterogeneity from unanticipated sources may emerge as important characteristics of the sample actually obtained.

Limitations in the number of patients falling within treatment groups in a study may sometimes make it impossible to show the significance of the difference between the treatments in that study. When, as is the usual case in Phase III investigations, each study is one of a series conducted under a common protocol, it is possible to combine the data from two or more such studies within a series to provide enough degrees of freedom to support a test of significance for the magnitude of the effect obtained.

The heterogeneity of the data may be so great that, despite a consistent trend in the effect of treatment and an ample number of degrees of freedom available for a test of significance, conventional criteria for statistical significance cannot be met. Some of this heterogeneity may be anticipated before the studies are undertaken and data analyses planned to reduce the heterogeneity. Often the heterogeneity cannot be anticipated, and its sources are not recognized until the studies have been completed. In such instances post hoc plans for data analyses to reduce heterogeneity may be applied.

In most Phase III investigations heterogeneity, whether anticipated or not, may be classified into three major areas:

a. Within a series there may be important heterogeneity from study to study. This may represent differences in treatment setting, source of referrals, attitudes and experience of investigators, etc.

   (1) Sometimes such heterogeneity between studies leads to the conclusion that the treatment is effective with certain kinds of patients and not with others and as a consequence results in more precise therapeutic applications.

   (2) When despite favorable trends statistical significance cannot be shown for each study separately, it is usually desirable to seek statistical significance by combining the data from the various independent studies (assuming a common research protocol). In such an instance it may be necessary to control for the between study heterogeneity before statistical

significance can be shown for the contrasts between treatment effects, e.g., by a factorial analysis of covariance.

b. There are always irrelevant differences between the treatment groups within any given study. These are usually pretreatment differences and occasionally are so great that they are statistically significant despite the fact that the treatments were assigned at random from the available pool of patients.

(1) Such pretreatment bias can be in the same direction as the desired treatment effect and enhance the apparent effect. The bias can be in a different direction and be sufficiently strong to obscure, if not reverse, the treatment effect. Pretreatment bias can be corrected in the data analysis by using different scores, residual scores which correct post-treatment scores on the basis of their regression on pretreatment scores, or analysis of covariance.

(2) Sometimes differences between treatment groups emerge during the course of study. Although the confounding effects of these differences cannot be removed from any one study statistically, these effects may appear to be more or less randomly distributed between the treatment groups from study to study within the series. Under such circumstances it may be possible to control for their confounding effects by introducing the confounding influence as a factor in the analysis of data combined from various studies within the series.

c. The magnitude of the heterogeneity within treatment groups in any given study may emerge as an important consideration.

(1) The greater the heterogeneity the poorer the chance of getting a significant difference between the treatment groups.

(2) If significant differences between the treatment groups is found despite great heterogeneity, two alternative interpretations should be considered:

(a) It is possible that the treatment is efficacious for all the varieties of patients included in the heterogeneous sample.

(b) It is possible that the treatment is highly efficacious for certain components of the treatment group but not efficacious for others.

(3) When the significance of the difference between two treatment groups is obscured by heterogeneity, the uncertainty can often be resolved in the data analysis if both of two conditions can be met:

(a) If the factor responsible for the heterogeneity can be identified.

(b) If the number of patients in each treatment group is sufficiently large to permit the introduction of the relevant source or sources of heterogeneity as a factor in the data analysis.

Sometimes such relevant diversities among patients can be anticipated on the basis of the Phase II studies, and the Phase III studies can then be designed so that heterogeneity is introduced systematically in the planning of the research. Often the presence and pertinence of relevant heterogeneity in Phase III studies is perceived ex post facto. When such sources of heterogeneity are introduced ex post facto into the data analysis to generate a significant difference, such "significant" differences must be viewed as a possible adventitious finding until they are confirmed by similar findings in at least one other independent study in the series.

(4) When within study sources of heterogeneity, not explicitly identified in the protocol as factors for statistical control, are introduced on an ex post facto basis into the analysis of data combined from several studies, a conclusion of efficacy must be based on some independent replication. When the series contains a sufficient number of independent studies, these studies may be combined into two or more groups, and the same post hoc control factor may be introduced into the combined analyses conducted independently for each of the two or more groups of studies. When comparable indications of significant efficacy are found for two independent groups of studies within a series having a common protocol, the possibility of an adventitious effect is greatly diminished.

In view of the types of analyses that may be required to handle problems accruing from heterogeneity and in view of the unfavorable discrepancy that usually emerges between size of the sample sought and the size of the sample completing the trials, some suggestions are offered for sample size. On the basis of experience, it is proposed that in most situations it would be optimal if 30 patients in each treatment group would have completed the requirements of the trials. Treatment groups of this size permit within study analyses of various sources of heterogeneity and can produce useful leads to identification of patients who may have responded especially well or not so well to the assigned treatment. In most instances as few as 20 cases completing the requirements of the study within each treatment group will be sufficient. With treatment groups of this size, however, the factorial handling of sources of heterogeneity within a study may be embarrassed because of a paucity of degrees of freedom. When fewer than 20 cases are available for each treatment group, the investigator may have to base his claim for efficacy on the results of the analyses of data combined from more than one study.

These general suggestions with respect to desirable sample size are only rough guides as the required size of the sample must depend on the magnitude of the effect and the magnitude of the sources of pertinent heterogeneity, quantities which can never be anticipated with precision.

8. The Number of Studies

The number of studies required for a series to support a claim of efficacy in some definable class of treatment setting and for some specifiable population is a function of the strength and consistency of the trends provided by the studies comprising the series. The somewhat circular nature of this statement reflects the fact that to sustain a claim of efficacy the studies must provide a basis for confident therapeutic application. This should require that all claims for efficacy must be confirmed by at least three strong, well conducted independent studies. (A claim of efficacy for some special subgroup could rest on two studies if both provided strong mutually confirming studies of efficacy and if the special subgroup were a part of some larger population or set of related subgroups for which strong independent evidence of efficacy already existed.) Specifically, if an investigator wished to base his claim of efficacy on a series of three independent studies, he should require that all studies show significant contrasts between the investigational compound and the placebo control for all of the criteria for which the claim is made. He should expect also to have at least 20 patients in each treatment group and to scrutinize these data in all samples independently by means of a common analytical procedure to identify the portions of the sample that were responsive and not responsive to the investigational compound. Under no conditions may he claim an efficacy for patients other than the kind shown to have been responsive in his samples.

When the criteria for which efficacy is to be claimed do not show a statistically significant advantage over placebo for the investigational substance in each of the first three studies comprising the series, the series must be extended to include more studies. For a series comprising four or more studies, it is suggested that a statistically significant advantage should be found for all the claimed criteria in at least one-half of the studies comprising the series. In addition, statistically significant advantage should be demonstrable for the claimed criteria for a combined analysis based on the pooled data from those studies which did not show the required significance when analyzed separately. It is assumed that if the protocol has not been violated all studies would offer treatment-placebo group contrasts in a direction to favor the hypothesis of efficacy.

C. METHODS

1. Frequency of Assessment

Some conditions tend to be episodic phenomena which, in most instances, remit spontaneously and often within a few weeks. As a consequence, the investigator must remember that he is examining the effects of a potential therapeutic agent in a changing context of pathology. The changes occurring spontaneously or in correspondence with factors extraneous to treatment tend to be of the same nature as changes sought by pharmacotherapy. For this reason, a potential therapy should be viewed from the standpoint of the speed with which it effects the changes, as well as from the level of remission eventually obtained. Since samples, as well as patients, differ from the standpoint of the severity of the problem, the duration of episode at the time that treatment is undertaken, the age of onset, developmental status of the child and resistance of the episode to treatment, one cannot be confident when appreciable spontaneous changes will appear. The investigator must be prepared, therefore, to make frequent assessments of changes in the patients. Investigators should not be reluctant to assess inpatients as often as once a day for some of new new drugs that are now in the process of preparation for large scale (Phase III) clinical trials.

2. Controls

Since the untreated course of most behavorial disorders cannot be specified in any standard manner, it is necessary that the effect of the compound under investigation be compared with the effect of placebo medication and possibly some standard medication, as well. The conditions of comparison should be as nearly equivalent as possible. This problem is ordinarily handled by assigning the patients to the alternative medications in some unbiased manner, usually random, and making the conditions of assignment and treatment as nearly double blind (i.e., unknown to both the patient and all members of the investigative team) as possible.

Bias can creep in after the medication has been assigned. For example, if a patient is not improving, he may receive more psychotherapeutic effort on the part of the treatment staff than the patient who is improving. A patient not improving may receive concurrent medication, such as sedatives, which under the protocol may or may not be allowable during the first few days of treatment, or he may receive other psychotropics, or other treatments, either assigned by the treating physician or self-assigned by the patient from other sources. Bias can develop also when the treatment is considered unsuccessful and the patient is withdrawn from the medication. Because of the various pressures that are placed on the physician treating the patient, it is important for the drug house monitor to be in frequent conversations with him and to examine the emerging data as they are being generated. In this way, the morale of the treating staff can be maintained and the probability of their following the conditions of treatment strengthened.

In Section B, Documentation, paragraph eight was concerned with the problem of showing significant contrasts between the investigational compound and placebo. Such contrasts are necessary to support a claim of therapeutic effect. Comparisons between investigational compounds and the standard medication are also required, not to establish efficacy, but to place the investigational compound relative to a treatment that is currently in accepted use for the kind of population, criteria of efficacy, and the treatment setting for which efficacy is claimed for the investigational drug. Obviously the efficacy of the new compound does not depend on its being superior or inferior to the standard medication.

3. Washout

One possible source of variability is the difference in the kinds of medication the patient may have had before he was assigned to the investigational program. If the prior medication had some therapeutic benefit and was still present in the blood stream to some appreciable degree, contrasts between the assigned treatment groups might be obscured, particularly if the patients (including the placebo group) had actually obtained and were obtaining some therapeutic benefit from a prior assignment of medication within the current episode. In addition, it may be feared that differences in prior medication might be an irrelevant source of variability within the sample.

If the disorder were of a chronic unremitting nature, the possible confounding effects of prior medication should be eliminated by a suitable lengthy washout period. In disorders which tend to occur as episodes with relatively short courses, a sufficient washout period (e.g., of one, two or more weeks) would result in a situation where any possible contrast between the effects of the drugs eventually assigned would be reduced by the spontaneously remissive changes that had already accrued. Therefore, patients taking drugs with a long half life such as phenothiazines, may be inappropriate for inclusion in other drug trials.

The half-life of a single dose of most psychotropic substances is known to vary substantially according to the drug, but the length of a washout period required to reduce the blood level of a psychoactive drug below the therapeutic threshold is not known. The duration of the prior medication and the size of the daily dosage are suspected to be important factors in determining the time required for a diminution below the level of therapeutic significance.

The importance of a preliminary washout period cannot be asserted with confidence, and investigators will not agree on this issue. In planning a washout period, it is important to consider several principles:

(a) The prior medication was probably ineffective; otherwise the patient would not be available for random reassignment.

(b) If the prior medication was effective there is no reason to assume that the therapeutic effect would invariably be undone by a washout period.

(c) An effective washout period could vary from a few hours through several weeks according to the medication. For this reason, a standard washout period could be expected to introduce variability in the duration of the period in which the patient has been without an effective quantity of medication, i.e., the unmedicated interval would vary.

(d) Most of the patients will be making spontaneous remissive improvement. After a few weeks of washout efficacy relative to placebo would be difficult, if not impossible, to show under most conditions of study.

(e) Compounds that are suspected of potentiating, nullifying, or in some other way modifying their effects interactively will have their own special washout requirements. At one time, accidents resulting from contiguous or concurrent use of amine oxidase inhibitors and tricyclic antidepressants provided a disturbing illustration of this point. The value of a washout period cannot always be assumed, however, and in any given case the ethical and scientific basis for the washout should be known and recorded.

4. Duration of Trials

Since no claim unsupported by clinical trial data can be made in the NDA or the package insert, this principle must be reflected in plans for the duration of trials. If the medication is being tested for its ability to reduce symptoms in patients who are resistent to other treatments, it would be reasonable to plan for clinical trials of 10 to 12 weeks duration. The current interest appears to be satisfied at present by a treatment period of four weeks. For fast acting drugs, trials longer than two weeks may become of little interest, and assessments at weekly intervals may come to be regarded as insensitive because of their infrequency. For most compounds at least one or two long term studies would be required in order to explore the possibility of untoward effects which could accrue from continued usage. Such trials should provide sequential assessments based on laboratory tests, vital signs, the usual side effects, and, of course, therapists' comments on any unexpected developments. For drugs in this area of use, there is also interest in the possibility of an habituation in the sense that increasing amounts of medication may be required to attain a therapeutic effect, or dependence in the sense either that a psychological demand independent of therapeutic need may emerge or that there may be significant withdrawal reactions if treatment is terminated suddenly after a long period of medication.

A presently undeveloped but potentially important area of treatment is the prophylactic use of relatively low dosage medication during remission for patients known to be subject to frequent repeated episodes. Such trials could be planned for a year or more, but the study of cumulative untoward effects, as well as confounding therapies and other detractions, would obviously be of pertinence in such trials.

5. Dosage Considerations

Unfortunately, many drugs come to the massive Phase III clinical inquiry before the effective dosage ranges for various indications and clinical subtypes have been explored. In such cases, it may be desirable for the relative efficacy of two or more alternative dosage levels to be explored, preferably in the same study, if necessary in parallel studies.

If the procedure provides for individual titration of medication on the basis of the investigator's judgment, it is important that the accompanying medical record include not only frequent (e.g., daily) notations on dosage requirements, but also notes specifying the reasons for dosage change and the nature of the patient's response. Such standard notations can be particularly useful for documenting the dosage recommended under various conditions. To increase the pertinence of the conclusions and to sharpen the evidence of efficacy, it may be possible in the data analysis to eliminate the information based on ineffective dosage levels and to base the claims for efficacy and recommendations for dosage on the analysis of data where the dosage level was appropriate.

6. Personnel

The choice of data gathering personnel must, of necessity, reflect the requirements of the assessment devices employed, the clinical setting under which data are

gathered, and the conditions of the patients. Insofar as possible, the data-gathering personnel should have the kind of training and orientation suitable for persons ordinarily involved in the treatment procedures. Ideally, persons involved in the treatment would also be involved in the data gathering.

Good initial training is not enough. Some provision must be made for continuing supervision of the data gatherers' efforts. Usually the best way of doing this is to arrange for a program where the emerging data are scrutinized at frequent intervals and any problems perceived are promptly shared with the persons involved. Often the pre-trial training of the data-gathering personnel can be combined with a pretesting of the applicability of the procedures. Certainly a pretesting of procedures is no less than prudent and does much to assure the pertinence of the procedures and the reasonableness of the conditions under which they are applied. Such pretesting should include the scoring, collating of the results, and scrutiny from the standpoint of both procedures of assessment and hazards to the protocol which emerge under the conditions of the study.

The number of persons involved in the use of any one assessment should be as few as possible within the study, and it is particularly important that, for any given patient, the persons who are involved in a given pretreatment assessment continue to be responsible for those assessment procedures throughout the period of treatment. In this way, heterogeneity accruing from the medication per se will not be confounded with changes due to differences in data-gathering personnel.

7. Departures from the Protocol

Violations of the protocol may reflect an unfortunate choice of clinician-investigators, an intrinsic difficulty in maintaining a standard plan with certain kinds of patient material or in certain treatment settings, faults in the initial design of the protocol, or faults in the training or supervision of the clinician-investigator or his assistants.

Concurrent treatments represent a most common departure from the protocol. Some of them, such as other psychotropics, and possibly psychotherapy, disqualify the case from further consideration, and it may not be known whether the introduction of concurrent treatment should be regarded as a treatment failure or as a failure of the clinician-investigator. Concurrent treatments for presumably unrelated disorders may also have a relevance for the symptoms under consideration. Examples might include concurrent oral antihistamines which may have a phenothiazine-like effect, thyrotropic medications, and the parent or patient's resort to his own personal supply of psychotropic substances. Such instances as these all comprise significant violations of the protocol and are cause for rejection of the case in question.

In otherwise properly conducted studies, it may be apparent that the protocol was not followed for one or two patients; sometimes the departure from the protocol is only accidental. It would be appropriate to eliminate such cases from the data analysis without impugning the rest of the study. In cases where some of the criteria could not be applied, it may be appropriate to analyze only the criterion data available; such situations should be scrutinized from the standpoint of biasing influences. In some instances, departures from the protocol may represent treatment failures in the sense that the patient's condition has worsened so that some other treatment had to be applied; untoward reactions may have occurred so that treatment had to be interrupted. Because such departures from the protocol are treatment-related, it is important to compare the treatment groups in terms of the number of cases where some departure from the protocol occurred.

Cases that have left the study are of particular interest. Sometimes this may represent a spontaneous improvement. Conceivably some of these early improvements may represent a rapid drug effect. Accordingly, the incidence of

early improvements may be used in the comparative evaluations of the medications in the study. Dropouts after the initial week may reflect a worsening of the patient and imply a treatment failure and the dropouts also should be analyzed as a criteria of efficacy. In other instances, departure from treatment may represent a medical or personal emergency which has nothing to do with the initial complaint of the present treatment. Often when a patient drops out of the study it is because of personal dissatisfaction on the part of the patient or his family, and the reasons for such voluntary withdrawal may never be known. It should not be assumed that they are necessarily treatment-related, but it is important to record the reason for dropout.

8. Comments on the Preparation of the Protocol

Although the content of the Guidelines is offered as having some general relevance for the preparation of the protocol, there are some special topics that should be considered, but do not fall readily in any one of the major categories.

One such area of interest concerns the criteria for excluding cases from the data analysis. Among the obviously unsuitable cases are those for whom the criteria reveal no pretreatment pathology. Unsuitable also are those individuals whose original diagnosis was in error and who were later perceived as not belonging in the population for which the efficacy claim is made. There are violations of the protocol which are sufficient cause for rejecting a case. Among the more serious of these is the confounding of treatment. It is particularly desirable that the protocol be unambiguous with respect to the kind and amount of confounding medication which would be sufficient for rejection. For patients with complex problems, it is particularly important to define the amount and kind of psychotherapy that is unacceptable. Other reasons for rejection include medical complications which can interfere with the treatment, obscure therapeutic effects, or invalidate the use of certain criteria. Faulty use of the criterion measurements can also be reason for rejection of the case.

Finally, there is the issue of the treatment resistant patient. These may be as important as the placebo responder who is in remission after the first or second day of a treatment which ordinarily takes one or two weeks to be effective. Since a post hoc formulation of criteria for rejection exposes the data to biasing selective influences, it is important that the criteria for the rejection be specified in the protocol and applied before the code which identifies the treatments is broken.

There are also reasons for excluding entire studies, and these, too, should be specified in advance. Among such reasons for rejecting a study in a pretreatment level of pathology which is shown by an important criterion or criteria to be so low that no significant diminution would be possible for samples of the obtained size and heterogeneity. Another reason for rejecting a study is a level of post-treatment pathology in the placebo group which is as low as could reasonably be expected for an effective treatment. Studies should also be rejected if the sample comprises an improper patient group, if the administration of the treatment does not follow the protocol, if the treatment setting is inappropriate, or if recommended procedures have been violated in the assessments.

The investigator is free to interpret criteria for rejection in any manner which seems appropriate for his studies, but he must specify his interpretation of these matters in advance and include them with the protocol. It is useful for the investigator to bear in mind that most violations of the protocol tend to obscure the differences that he is trying to show, but if his study reveals the required significant advantages for the investigational compound despite such violations of the protocol, rejection should not be automatic or necessary.

The protocol should also specify who the members of the investigational team should be, how they are recruited and trained, and what roles they should play and under what conditions. The selection and preparation of investigators for outpatient studies can be particularly critical, and it is important that the overall protocol statement include instructions concerning the orientation of the patient to the treatment. Specifically, the doctor should be interested in learning about all the kinds of medication the patient has recently used and about the patient's family's supply of unused medication (particularly sedatives, tranquilizers, and antidepressants) from former perscriptions.

The patient should be cautioned not to use his private supply of medication without first consulting the doctor, and the patient should be urged to report the use of any unauthorized medication promptly.

9. Analytical Considerations

Since much has been written elsewhere concerning the analysis of clinical trial data, the present comments are limited to two considerations which are commonly disregarded in clinical trials.

(a) The data from most clinical trials is amenable to two somewhat different analytical approaches, qualitative and quantitative. The qualitative enumerative procedure is based on counts of various categories or levels of patient status and change. The comparisons are then expressed in terms of the frequency of various kinds of responses or levels of ratings or in terms of the frequency with which positive or negative changes occur, etc. Data in this form lend themselves to the use of nonparametric procedures. The analyses of data reflecting such a qualitative, enumerative approach are usually relatively simple, although more complex analyses involving factoral considerations are possible. An enumerative approach to the data can illuminate weaknesses or reveal strengths which are not apparent in the more familiar quantitative treatments.

Although inventory items and discrete symptom rating scales tend to involve a small number of alternatives and lend themselves to enumerative analyses, they usually imply a continuum and are considered suitable for quantitative analysis.

The quantitative treatment of the data appears to be the preferred analytical approach but it should not be exclusively. The quantitative approach emphasizes means and variances, lends itself to complex multivariate designs, and is oriented toward parametric tests of significance. Many of the criteria used in clinical trials are composite scores often based on prior factor analyses of symptom rating scale items or inventory items. These composite scores usually comprise many steps and permit the direct application of quantitative approaches. For the practical application of enumerative procedures, these composite scores would have to be submitted to arbitrary disjunctions; such a procedure would obscure discrimination and amount to a diminution of the sensitivity of the measures.

Thus, the quantitative parametric approach is most desirable for continua involving numerous steps, particularly composite scores. For the simple continua represented by the small number of alternatives provided by most inventory items and discrete symptom rating scales, either a quantitative or a qualitative enumerative approach is applicable. When the alternatives do not generate any conceivable continuum, the quantitative approach is inapplicable, and only a qualitatively oriented enumerative analysis is possible.

Wherever reasonable, both a qualitative and a quantitative approach to handling of the data should be provided. Qualitative analyses have the advantage of indicating the portion of the sample involved in a change, while quantitative

analyses show the average level of change with no indication of the portion of the sample involved.

(b) Since, by their nature, self-descriptive inventories and symptom rating scales comprise numerous items, the data analysis, in effect, involves many criteria. For those inventories or sets of rating scales where some standard provision is made for combining the items into scores, some clinical trials involve comparisons based on several such composite scores, sometimes as many as ten or more. In any situation where treatment groups are compared in terms of numerous criteria, some scores will emerge as statistically significant as a result of purely adventitious factors which have no relation to the treatment effect per se. Two types of considerations may be applied to the problem of deciding whether the results have a systematic association with treatment or whether they can be ascribed to chance. If the sample is large enough to supply a sufficient number of degrees of freedom or if the number of criterion measures is small enough so that the requirements for the degrees of freedom are not excessive, a multivariate discriminant analysis may be applied to answer the question of whether the set of variables significantly discriminates between the treatment groups. Testing the discriminatory significance of a set of criteria disregards any qualitative distinctions among the criteria, whether the distinctions are a priori or post hoc.

Sometimes in a series of multivariate studies under a common protocol it is apparent that certain measures never discriminate between the treatment groups and other variables do discriminate in at least some of the samples provided by the several investigators. Under such circumstances, it would seem witless to apply blindly a multivariate discriminant analysis to all the criteria, including those which never discriminate.

D. CRITERIA

The criteria for the assessment of changes during the course of clinical trials should be selected in relation to at least five different kinds of consideration:

1. Criteria should be selected in relation to the indications for which it was hoped, if not expected, that the experimental medication would be efficacious. This would mean that the criteria must be pertinent in two respects:

   (a) The criteria must reflect the kinds of symptoms, subjective discomforts, impairments of performance, and possible psychosomatic equivalents for which evidence of efficacy is desired.

   (b) The criteria must provide distinctions at a level of psychopathological disturbance which is expected to be modified by the treatment under test. For example, it is possible to show changes in anxiety among moderately uncomfortable patients by the use of devices which are not appropriate for showing changes in the panic-like disturbance frequently found in fulminating psychoses. In contrast, measures of anxiety which may be appropriate for revealing changes in psychotic patients may be insensitive to changes in the kind of anxieties commonly encountered in outpatient psychiatric practice. Thus, the level of pathological disturbance or impairment for which the medication is expected to be efficacious should be considered to assure that the criteria can reveal changes at a relevant level of severity.

2. The criteria should be selected in relation to the kinds of patients for whom efficacy may ultimately be claimed. For example, if the medication will be tested with literate private outpatient adolescents, self-administering inventories which represent the subjectively experienced distress of the patient may be sensitive to the

level of changes desired. Such patients must be able to recognize their current discomforts in the items of the inventory and not confuse an assessment of their immediate state with an evaluation of their enduring traits.

The appropriateness of a self-report inventory may also depend upon the sophistication of the patient. For example, a psychologically naive patient or a repressive patient may recognize his distress in somatic equivalents or in performance loss only. He may be unprepared to acknowledge the subjective uncertainty and dread that is apparent to the psychiatric rater.

If evidence of efficacy were desired for patients whose only indication was subjective distress, but who were illiterate adults or children, the usual self-administrating inventories would be inapplicable unless they could be administered in an interview type situation which would, in effect, require that the interviewer interrogate the patient with respect to each of the inventory questions. This kind of involvement with an interviewer imposes some requirement for the training of the person who administers the inventory.

Among many patients, the subjective discomfort ordinarily measured by an appropriate inventory is accompanied by symptomatic developments which are revealed in the course of the interview and may be rated by an appropriate rating scale. Rating scales have been found to be more sensitive to treatment effect than self-report inventories.

Patients with a variety of psychological deficits often show some impairment of performance in areas not directly related to their target symptoms. This impairment of performances may be reflected in various practical respects. As interference with general performance increases, the patient's problems are more and more likely to be reflected in restricted performance in various kinds of situations calling for cooperation, participation, or general interest in surrounding events. For this purpose, a comprehensive symptom rating scale or a behavior rating scale is required. Thus, the level of severity at which the medication is directed has implications for the choice of assessment devices.

Since discomforts and symptoms rarely develop in isolation, but are usually a part of a pattern or syndrome of symptoms, the criteria must reflect not only the presenting cause for treatment, but also the accompanying complaints or symptoms. The spectrum of assessments, whether subjective, symptomatic, psychosomatic, or any combination of these, requires a somewhat different set of criteria depending on whether the patients are psychotic, rational but unsocialized, or able to function in age-appropriate settings but uncomfortable. It may also depend upon whether the accompanying symptoms involved severe learning problems, problems with authority, family disorganization, paranoid caution and rigidity, an obsessively ruminating preoccupation, severe acting out with the character implications, or shallow histrionic manipulations.

The investigator is free to examine any set of indications for which it is desired to claim an efficacy, but efficacy cannot be claimed for indications which are not represented in the criteria.

3. The criteria should be selected in relation to the conditions under which the drug will be tested and under which it will eventually be used. This consideration places substantial restrictions on the choice of criteria for large scale clinical trials. Regardless of their possible pertinence, many devices for reflecting drug effect, such as EEG tracings under various conditions, computer assisted continuous performance tests, speech samples to be analyzed according to certain dynamic principles, or according to the duration, latency, and other temporal characteristics of response as estimated in a standard interview or as evaluated by a special chronograph, may not be practical under the conditions of large-scale clinical trials.

Similarly, aspects of family, community, or school adjustment which require the participation of a data-gathering social worker are also impractical in most situations. Certain types of interesting laboratory tests of metabolic features are similarly impractical. These special criteria are applicable only where the necessary equipment and the appropriate personnel are available.

Even the simplest of performance tests should not be undertaken unless the conditions of testing can be standard and the personnel responsible for the testing aware of the importance of maintaining standard conditions and procedures from test to test.

Perhaps the most important single guarantee that the investigator can have concerning the appropriateness of the assessments results from pretesting the proposed assessment procedures with the kinds of patients desired for clinical trials and under the anticipated conditions of treatment. Such pretesting can be a part of the indoctrination and training of the clinician investigator and members of his staff who participate in the investigation.

If the concept of an actual pretesting of all the procedures seems alien, perhaps it should be construed to mean that the data from the first 10 patients would be regarded as a pilot study and be intensively reviewed by both the study monitor and the members of the clinical investigative team in joint discussion. As a result of these discussions, the techniques and procedures of the clinical investigative team, and perhaps the choice of the assessment devices, as well, would be modified in order to arrive at a procedure which would be both practicable under the conditions of the study and faithful to the purposes of the study.

4. From the standpoint of changes to be assessed, characteristics of the patients treated, and conditions of treatment, the choice of criteria can be guided by reference to the published literature. Much can be said for using assessment devices which have been found to be efficacious in revealing the expected changes in patients for whom it is hoped efficacy may be exhibited and under the anticipated conditions of treatment.

If the published literature provides no pertinent precedents, the study must be planned and the criteria must be selected without this source of reassuring guidance. Under these pioneering conditions where there is no precedent, it is particularly important that the investigators select criterion devices which can be reliably used under the conditions of treatment. The most important assurance can be provided by small pilot studies which are identified with the FDA as exercises for the selection of appropriate procedures and not as data which will be presented in support of a claim of efficacy. In any pretesting or pilot study, it must be remembered that if a criterion does not show an appreciable pretreatment level of a symptom or complaint, that condition cannot show an improvement during the course of treatment. Moreover, the presence of patients whose pretreatment pathology is not indicated by the criteria can obscure therapeutic effects despite the presence in the sample of other patients whose pathological indications are known to be remitting. Whether the difficulty lies in the choice of criteria or in the selection of patient material, pretesting or pilot procedures can protect the study from a major source of insensitivity.

In planning studies, it is important to give a high priority to the type of information desired. For inventories to be sensitive to change in the current state of the patient, the content of the inventory items must refer to qualities for which change is desired. It is not unusual for criterion measures to be selected with greater consideration for their familiarity and general acceptance than for their pertinence to the goals of the inquiry. A relevant ad hoc assessment device may prove of much greater value than an esteemed traditional but somewhat inappropriate assessment. One should not underestimate the hazard that a substance of substantial therapeutic

merit may be rejected as ineffective because it was tested with inappropriate criteria, with an inappropriate patient population, under inappropriate conditions, or at an ineffective dosage level.

During the lifetime of the present guideline statements, blood levels of the medication and perhaps some toxicity tests as well may be considered necessary to provide documentation of the medication as the agent for the therapeutic changes and in this sense may be regarded as criteria for the availability of the therapeutic agent. Standard methods to implement this interest are not yet available, but they are generally recognized as desirable, particularly for studies of efficacy in chronic administration.

5. Criteria should be selected with a concern for their reliability. The reliability of assessments accrues from many sources, including the nature of the patient material, the treatment setting, the morale of personnel, and the training and supervision of personnel. The content and the construction of assessment procedures have certain inherent amgibuities so that the variability of the scores is inflated to some degree by uncertainties and inconsistencies intrinsic to the assessment device per se.

   Accordingly, under the most optimal circumstances, some criteria have important limitations in their discriminating potential. Although instruments which have been repeatedly used with success in drug trials may sometimes be assumed to have at least a useful degree of reliability, reliability cannot be assumed for untried assessment devices.

   Regardless of the assessment device selected, the reliability of the scores obtained in a given circumstance is somewhat dependent upon those circumstances. One cannot assume that the level of reliability demonstrable under favorable circumstances will be present under all circumstances of use. Pilot testing can reveal inconsistencies and lead to refinements which strengthen the reliability. A useful discussion of the reliability, objectivity validity of clinical rating scales may be found in the Journal of Nervous and Mental Disease (Vol. 154 No. 2, 1972, pp. 79-87).

E. CLINICAL SIGNIFICANCE

It is conventional to compare a therapeutic response to the medication in question with therapeutic response to some control substance, e.g., placebo. The comparison is expressed in terms of summarizing statistics, such as mean differences or differences in the portion of the respective samples meeting a certain criterion for improvement. It is well recognized that the direction of such contrasts is not necessarily a sufficient basis for conceding or denying efficacy. Certainly the differences between the two treatments under comparison must be greater than could be explained as a random fluctuation which might occur by chance alone under the conditions of the study. This question of statistical significance may be answered in terms of any one of several conventions. It must be remembered, however, that level of statistical significance may accrue from the homogeneity of the sample or the size of the sample, as well as from the magnitude of the contrasts between the trends in the two samples. When the two samples are large and the heterogeneity within the samples is quite small, a clinically trivial difference can meet a test of statistical significance.

Before clinical significance may be ascribed to the various statistically significant differences which may be found in the course of the various studies required by the Phase III clinical trials in support of a new drug application, several factors must be considered.

"Clinical significance" is a judgement of the meaning of the obtained "statistical significance;" a statistically significant difference obtained between two treatments has no necessary clinical meaning per se. Although it is obvious that a clinically significant

difference will meet ordinary criteria for statistical significance, a statement of statistical significance alone conveys no assurance of therapeutic value. Clinical significance accrues from the meaning that may be ascribed to the difference of differences which have met the statistical criteria and accrues from context.

The most meaningful context for a given clinical significance is provided by the circumstances under which the treatment will be used. There are two classes of circumstances that may be considered. One class is broadly general and refers to the conditions that can be expected to accompany the indications being treated. The other is specific to the individual patient and must be deduced by the physician responsible for treating the individual patient.

Clinical significance is a judgment made by the user of the medication, and the nature of the judgement is based on the confidence with which the user anticipates the outcome of the treatment. For this reason, clinical significance remains unknown until some estimation of the portion of the sample which meets the level of improvement for which efficacy is claimed. This estimate must be provided for patients treated with the investigational substance, for patients treated with the standard substance, and for patients treated with placebo. For example, if, after treatment, the level of severity in the placebo group were reduced to such a degree that 60% of the patients were in the normal range, it would be reassuring to know that 90% of the patients treated with the investigational compound had a level of severity within the normal range. For many criteria, the normal range is not satisfactorily defined, but various alternatives suggest themselves, for example, it may be possible to compare the two groups with respect to the portion of patients who, at the time of post-treatment assessment, had a level of pathology which would place them in the lower 10% of the pre-treatment distribution.

The studies should examine characteristics of those patients in whom an appreciable response was found and of the characteristics of patients in whom a response was not found. Such information did increase the confidence with which the medication may be assigned to a given patient and confers additional clinical significance to the treatment.

In addition, the studies should compare the speed with which remission or amelioration occurs in the treated group with the speed with which such ameliorative changes occur spontaneously in patients treated with placebo and in patients treated with some alternative medication.

Clinical significance must consider the "cost" factors associated with the use of the medication in question. The "cost" must be reckoned primarily in terms of risk and inconvenience to the patient and secondarily in terms of monitary and personnel considerations.

To estimate the "cost" of the treatment, the therapist must be informed of the nature of any untoward reactions, the frequency with which they may be expected in various identifiable classes of patients, and the nature of the management problem that these untoward reactions may generate. The therapist must also be informed of the risk of habituation which may be in the form of a physiologic adaptation with the possible requirement of increasing dosage level and possible withdrawal problems. If there is risk of a withdrawal problem, the clinical significance of a therapeutic claim must also involve information concerning the management of this aspect of habituation. In addition, there is the possible risk of psychological habituation. The probability of such an acquired demand must be assessed realistically and included as a part of the basis for a claim of clinical significance.

It is not admissible to claim or imply a therapeutic benefit for any group of patients or for any quality of pathology not directly represented in the sample effectively treated or by the criteria which consistently distinguish between treatment groups. For example, if global improvement as rated by psychiatrists is the only criterion which distinguishes between the two treatment groups with consistency and statistical significance, it is

defensible to claim only that in the opinion of the treating psychiatrist the patients were improved relative to placebo. By their very nature global improvement ratings have a basis which can vary from case to case and for this reason their basis for a sample of patients is unknown. Accordingly, global improvement ratings proved no justification for claims of improvement concerning such detailed matters as how the patient feels, the diminution of any symptoms, the patient's work performance, or his ward behavior. Similarly, if consistent improvement in certain symptomatic respects is the only supporting criterion, it cannot be claimed or implied that the patients are generally improved, that they feel better, etc. Similarly, if only the total score or some part score of a self-descriptive inventory discriminated between the investigational group and the control group, claims for only this aspect of therapeutic benefit may be claimed, and claims for other therapeutic benefits should not be stated or implied.

APPENDIX II

GENERAL CONSIDERATIONS FOR THE CLINICAL EVALUATION OF DRUGS IN INFANTS AND CHILDREN

I. GENERAL PRINCIPLES

A. INTRODUCTION AND OVERVIEW

The booklet entitled "General Considerations for the Clinical Evaluation of Drugs" contains much information which is applicable to drug testing in children and it should be considered a companion piece to this booklet.

To facilitate approval of new drugs for use in children testing should be related to the anticipated duration of usage and to the size and age of the pediatric population likely to be exposed to the new drug. Emphasis should be placed on elucidation of unexpected toxicity, not simply collecting examples of the types of toxicity predictable from knowledge of the pharmacologic properties of the drug. New and innovative forms of in vitro and in vivo testing should be employed because new agents developed today, which may exhibit some of the same forms of toxicity responsible for therapeutic catastrophies of the past, may not be identified as such by current testing procedures.

The design of studies must be flexible to recognize the need for evaluation of a new drug or substance for the treatment of rare diseases or diseases which are unique to the pediatric age group. In these circumstances, special considerations may include an abridgement of the usual requirements for safety and efficacy. Such abridgement should be considered when the use of the drug is limited to a few patients, particularly patients suffering from a disease for which no alternate therapy is available. In addition, an investigator concerned with such patients should be allowed considerable latitude to administer various substances, particularly naturally occuring amino acids, cofactors, and vitamins without extensive preclinical studies. Furthermore, if no appropriate animal model for a disease condition exists, and if efficacy is readily demonstrable (e.g. certain seizure patients), early efficacy studies in children are appropriate.

B. FACTORS AFFECTING BOTH SAFETY AND EFFICACY

1. Methods

Adequate methods for determination of the drug and its major metabolites (especially those which are pharmacologically active) in biologic fluids (especially serum and optimally in tissues) should be developed during preclinical or early clinical (phase I and II) testing. The particular method obviously will depend on the chemical nature of the drug, expected concentrations in serum, etc., but it should not require administration of radiation emitting substances. Assays based on techniques such as radioimmunoassay, gas-liquid chromatography, and competitive protein binding are at present the most likely to achieve the desired degree of accuracy, sensitivity, and reproducibility. Use of stable isotopes is a method of great promise, although the initial cost of equipment may be prohibitive except in research centers and the National Center for Toxicologic Research. The administration of radioisotopes to children is not to be generally condemned, but it should be avoided except under special circumstances. Such techniques are of great value and entirely appropriate for special studies under appropriate circumstances. For example, use of tracer amounts of labeled (^{14}C, ^{3}H) amino acids, glucose,

or other intermediary metabolites may be invaluable for defining metabolic diseases, and similar employment of labeled drugs could conceivably be employed. Use of isotopes, other than ^{14}C and ^{3}H, which have short half-lives and low-energy emission equivalent to a conventional chest- x-ray offer considerable promise and should be employed whenever possible.

The small sample volume obtainable, particularly from small infants, is a critical factor in the development of appropriate methods, particularly when multiple samples are required. This is not a prohibitive requirement and should not be used as an excuse to avoid development of appropriate assay procedures. Radioimmunoassays for drugs such as digoxin or diphenylhydantoin have been developed which utilize as little as 20 to 100 microliters of serum. The development of appropriate methods for determination of serum levels is particularly important for those drugs in which serum levels can readily be related to pharmacologic or therapeutic effects. In these instances, determination of serum levels is the key to studies of dose, dose interval, bioavailability (when coupled with urinary excretion), apparent volume of distribution, etc.

Methods should be continually reviewed, revised, and updated with the goal of developing methods appropriate for routine use in laboratories cooperating with the investigator, and such assays should become sufficiently standardized and simplified so they are within the practical capability of the clinical laboratory of any large hospital. Moreover, modifications should be directed toward identification and quantification of the principle metabolites of the drug, so comparison may be made with the elimination pattern of adults. If major differences exist, such studies would serve as a warning of possible adverse effects and should lead to attempts to identify the unique pathway of metabolism in the immature patient.

With certain categories of drugs - the so-called "hit and run" agents, such as the cytotoxic drugs, certain enzyme inhibitors, storage granule depletors, etc.
for assay methodology may be relaxed or waived. Other appropriate assays of biologic effect should be developed for these agents. For example, inhibition of incorporation of tritiated thymidine into white blood cells might be used as a measure of the effect of certain cytotoxic agents. Antibiotics and certain other chemotherapeutic agents have special requirements and methods for estimation of effective serum levels. Bioassay techniques are entirely appropriate as long as the method is scaled down to the small sample volume of pediatric patients. Techniques employing the patient's own pathogen as the test organism should be available for the use of clinical laboratories engaged in phase II and III trials.

2. Studies of Absorption, Distribution, Metabolism and Excretion (ADME)

Studies with varying degrees of depth and completeness, appropriate to the drug and its intended use, are essential for each age group and are described in detail in the respective sections. In general, the preclinical and early clinical phases should lead to accumulation of data which account in a major way for the disposition of the drug. Not every metabolite may be identified, and the intimate details of each of the ADME phases will not be elucidated. Judgment must be exercised about requirements for data which are clinically relevant, and not all drugs should be subjected to full investigation. However,
the following data should be available for drugs which will be administered orally in divided doses for courses of one week or longer:

a. Absorption: From the physical nature of the drug and its pKa the influence of changes in pH of the stomach and intestine on the ionization and thus

the absorption of the drug can be predicted and verified. When appropriate, the approximate percentage of a single oral dose absorbed should be determined. If easily studied and when of possible clinical importance, the area of the gastrointestinal tract where the drug is absorbed (i.e., stomach, terminal ileum, etc.) may provide useful information in predicting drug interactions and alterations in absorption in disease states.

b. Distribution: Binding to plasma proteins (affinity and percent bound at therapeutic blood levels), whether albumin, globulins, or special carrier proteins, and the percent of total serum concentration which is "free" should be determined. Distribution and particular propensity for accumulation or fixation to certain tissues (for example, tetracycline in bone and teeth) in developing and mature animals should alert reviewers of possible forms of toxicity so appropriate additional studies can be requested. Apparent volume of distribution may be useful in designing dosage regimens. Studies of dialyzability may be useful in developing recommendations for the management of overdoses and accidental ingestions.

c. Metabolism: The pattern of metabolites and the biotransformation reactions involved - that is, hydroxylation, demethylation, glucuronidation, etc. - should be known from studies in man. Requirements for toxicity studies in immature animals (especially rodents) should be limited, if possible, to a species for which experimental evidence has established a similarity by immature humans to the handling of the agent being tested.

3. Bioavailability

An important influence on studies of safety and efficacy is the bioavilability of different formulations and of different manufacturers' products. When the dosage form constitutes a new chemical entity, appropriate studies must be conducted in adults before children are exposed. The exact and total constituents of the final dosage form should be known. Studies of bioavailability should include, but not be limited to, determination of serum levels and the time of peak levels after a single dose. Total absorption is usually best determined by quantitative determination of the urinary excretion of the drug and its principal metabolites. Because of differences in pH, gastric emptying time, intestinal motility, etc., differences in bioavailability, especially between newborn infants and adults, should be duly considered and investigated when appropriate. Moreover, when changes in gastric or intestinal pH, flora, or motility might be reasonably anticipated to differ from normal because of disease or other factors, additional studies are indicated. Studies of bioavailability often may be sufficiently covered in conjunction with studies of absorption, efficacy, etc., and need not demand independent investigations.

The possible toxicity or influence on the pharmacologic properties of the drug by the vehicle and/or other components of the formulation (stabilizers, excipients, etc.) must be considered. This results from the fact that many drugs tested in the form of tablets or capsules in adults will be administered as suspensions, solutions, or elixirs to infants and children. Moreover, the vehicle or solubilizing chemicals in parenteral preparations must be considered as a possible source of uniquely toxic agents, particularly for newborn infants.

4. Drug Interactions

Interactions between drugs occur in a variety of ways, ranging from physiochemical incompatibilities to opposing or synergistic pharmacologic effects. Preclinical and in vitro testing can be expected to detect most interactions,

particularly when coupled with phase I and II testing in adults. However, especially in neonates, age-dependent differences in pharmacokinetics may result in unique interactions. For appropriate review of a new agent, the types of drugs which may be used in conjunction with the proposed agent for the same disease or condition at different ages should be considered to completely evaluate possible drug interactions.

Physicochemical interactions will probably be detected in early work with the new drug. Of particular concern in pediatric usage would be interactions which might interfere with the absorption or action of vitamins, trace minerals, essential amino and fatty acids, or other constituents of infant formulas and other dietary sources.

Physiologic or pharmacologic actions which might further impair the normally limited capacity of the neonate to metabolize and/or excrete drugs would be of particular concern. Specifically, inhibition of or competition for hepatic biotransformation reactions occurring via the mixed-function oxidase system and/or the glucuronide conjugating system, or decreases in glomerular filtration rate or tubular secretion can be predicted to have important consequences for the newborn.

Further interactions of particular concern to newborn infants relate to bilirubin, particularly with drugs administered near term, at delivery, or directly to the newborn. Binding to albumin with displacement of bilirubin and enhanced neurotoxicity is known to occur with a number of anionic compounds. Other factors (e.g., hypoxemia and acidosis) have also been reported to increase the potential toxicity of bilirubin. Moreover, binding by drugs might interfere with the transport and action of endogenous substances other than bilirubin (cortisol, thyroxin, fatty acids, etc.) and with the binding of other drugs.

5. Enzyme Induction

The importance in pediatrics of the induction of hepatic drug-metabolizing enzyme activity by exposure to drugs and chemicals is unclear at present. Three hundred or more drugs and chemicals are known to produce marked increases in liver size, proliferation of smooth endoplasmic reticulum, and increases in the specific activity of mixed-function oxidase and glucuronyl transferase enzymes in experimental animals. In clinical studies, small changes in serum concentrations and half-life for a few drugs have been reported in adults, although some negative reports have appeared.

Almost nothing is known about "inducibility" at various ages in man. Decreases in serum bilirubin levels have been reported in congenital nonhemolytic jaundice and in normal infants with "physiologic" jaundice treated with phenobarbital, nikethamide, and DDT. Increased smooth endoplasmic reticulum in hepatocytes and increased NADPH cytochrome c reductase (a microsomal enzyme) activity have been shown in infants treated with phenobarbital. Similarly, increased glucuronidation of salicylamide has been reported. Thus, the infant can respond to exogenous "inducing" agents although the details of the process and the extent and the clinical importance of this reponse remain unclear.

When induction is considered relevant, noninvasive types of studies, such as antipyrine half-life as determined by salivary concentrations or urinary excretion of the hydroxylated metabolite, may be undertaken. The urinary excretion of 6-hydroxycortisol or D-glucaric acid may also be used as monitors. Invasive techniques - such as direct determination of serum half-life or, rarely, liver biopsy obtained adventitiously - may yield more direct data.

C. EFFICACY

Because of ethical considerations, reasonable evidence of efficacy generally should be known before infants and children are exposed to the agent. Testing against the best known agent will be the preferable method for establishing efficacy with many drugs. A drug may be useful for only a certain percentage of the population diagnosed as having a general broad category of disease. For example, it is entirely possible that only a relatively small percentage of the "disease" population with bronchial asthma (a disorder probably of multicausal etiology resulting in similar clinical manifestations) may benefit from a particular therapeutic agent. In contrast, evaluations of efficacy at times may deal with an extremely small population. For example, a useful agent might demonstrate efficacy after study in only a few patients with a rare aminoacidopathy. Therefore, the requirement for demonstration of efficacy must not deal with fixed numbers. Again, flexibility must underline decisions about the number of subjects in each phase.

Based on ethical considerations, sick children rather than well ones will be the principal source of the experimental population, therefore, placebo groups cannot always be employed. Obviously, therapy cannot be withheld or an inactive drug cannot be administered by injection or other painful procedure. A number of alternative methods to the classical double-blind placebo experimental design can be suggested. In many instances, a standard drug can be used for comparison. Historical group controls may be utilized. "No drug"-crossover can be used if the patient can tolerate a "no drug" period without serious compromise of his health. At times, the patient may serve as his own control, either as a personal historic control or in a "crossover drug/no drug" or "drug/standard drug" design. The drug may be most importantly compared to other therapeutic modalities, for example, behavioral modification, psychotherapy, dietary manipulation, and so forth.

Specific types of diseases where efficacy is likely to be tested are described for each age group in Section II.

D. EXPERIMENTAL DESIGN

Ethical, practical, and legal considerations may preclude studies by the most theoretically ideal experimental approach. This fact need not be viewed as an insurmountable obstacle because drugs should optimally be tested under conditions of actual clinical use, whether administered to hospitalized patients or in office practice. Such considerations do not obviate the need to establish a rigid protocol, including appropriate controls of whatever type, evaluating dose response phenomnena, and adhering to sound experimental design.

Study design must: (1) account for adequate control of variables and include appropriate statistical procedures, (2) detail methods and provide validation for assessment of benefit, (3) allow for handling of adverse or side effects, and (4) demonstrate awareness of the placebo response, both for beneficial and for adverse effects.

Perhaps the single most important variable to be assessed and controlled is the comparability of the study populations. This must be assessed in terms of a variety of parameters appropriate to the study, at times including but not limited to disease, social, physical, intellectual, and behavioral equivalence.

The mechanism(s) for evaluating adverse effects, whether by means of volunteered or elicited reports, questionnaires, or other means must be clearly stated and appropriate for the age group(s) under study.

Provision should be made for the management of accidental or intentional overdosage and severe, acute toxic reactions. Dialysability, specific antidotes, and other therapeutic measures should be assessed, and such information should be included in the

protocol which is available to all involved in the study.

There should be safeguards to ensure that any study can be terminated at the earliest possible moment if danger to the subjects arises.

Studies of blood, liver, and renal function should be selective and appropriate for known modes of action and toxicity, rather than the accumulation of a mass of laboratory data from samples obtained by venipuncture or other painful procedures which are then run through the autoanalyzer. Initially, a wide base of studies may be used; but, if these studies are negative, only a few highly selective parameters should be monitored. A similar approach is suggested for the use of ECG, EEG, and other time-consuming and expensive studies.

II. SPECIFIC AGE-DEPENDENT FACTORS INFLUENCING SAFETY AND EFFICACY

Growth from conception to adult life involves complex changes in anatomy, physiology, biochemistry, and behavior which vary considerably from one state of development to another. Therefore, the action and adverse actions of pharmacological agents will vary as absorption, distribution, metabolism and excretion, and receptor sensitivity are altered by the changes associated with growth and development.

In recognition of these developmental changes, this portion has been written in sections; periods of childhood have been divided into stages which share characteristics distinguishing each stage from the other stages. In each stage, factors which may influence the disposition and action of a drug and the major immediate, delayed, and adverse actions are related to the major biologic events of the stage.

By introducing these age groups, it is not suggested that each drug be tested in each age group; rather, this is an attempt to ensure that the important biologic characteristics of the age(s) in which the drug eventually will be used therapeutically will be considered in evaluating both its beneficial and its undesirable effects.

Each age group will be evaluated as follows:

1. A General Statement of the biochemical, physiologic, and behavioral characteristics of the age group; specific ways in which the child is unique at the stage will be given.

2. Safety Considerations of particular importance to the age group. These are divided into three subgroups relating to the type of toxicity encountered and the temporal relationship of these effects to the initiation of therapy.

   a. Immediate Toxicity: Signs and symptoms occur soon after the initiation of therapy.

   b. Delayed Toxicity: Toxic effects occur only after a period of chronic administration. Certain adverse effects which occur in the immediate period of administration but manifest themselves later (such as tetracycline staining of the teeth) are also included in this category.

   c. Late Onset Toxicity: Toxicity which becomes apparent months to years later, e.g., adenocarcinoma of the vagina in girls born to mothers who received diethylstilbestrol during pregnancy.

3. Efficacy

   Means of establishing the beneficial effects of a drug and particular forms of desirable therapeutic activities.

4. Problems in Drug Evaluation

Special problems which may arise in the evaluation of drug action in a given age group.

5. Ethical Considerations

Special ethical considerations pertinent to each age group are delineated.

A. INTRA-UTERINE (CONCEPTION TO BIRTH)

1. General

The administration of drugs to the pregnant woman presents a unique problem to the physician. He must consider maternal pharmacologic mechanisms, and he must be aware of the fetus as a recipient of the drug. In therapeutic endeavors directed toward maternal disease, consequences of drug usage have often been unexpected; and adverse effects have appeared in the developing fetus, for whom the drug was not intended. On the other hand, the possibility of development of drugs for the treatment of fetal disease diagnosed in utero should be considered, and guidelines should be developed for the evaluation of both efficacy and safety of this type of compound when it is administered either via the maternal route or directly to the fetus. Drugs may also be administered to women who are not aware they are pregnant.

2. Safety and Efficacy

Adverse effects of drugs on the fetus vary depending on the stage of intra-uterine development. Before implantation, drugs may appear in high concentrations in tubular fluid and lead to the death of the fertilized ovum. Drugs which cause an adverse effect during organogenesis may result in anatomic malformations. Drugs given beyond the period of organogensis may affect the fetus and cause a functional disorder which is not associated with any known anatomic malformation.

Suggested methods of procedure to evaluate drugs which may be given to the mother during intra-uterine development are given in the following paragraphs. A prerequisite to intra-uterine studies for any new drug is evaluation (phase I and II) in adult men and in nonpregnant women of childbearing age.

Organogenesis--To evaluate drugs which will be used in pregnant women during the period of organogenesis, pharmacokinetic studies should be conducted in animals, including a subhuman primate. Localization of the drug within the fetus may be readily accomplished using isotopic techniques. At the same time, although not mandatory, studies of drug metabolism and disposition within the human fetal-placental unit should be considered.

The next stage of intra-uterine development to be considered for drug evaluation is from the completion of organogenesis to the onset of labor. This separation from the other periods of intra-uterine life is arbitrary because there will be drugs used throughout pregnancy for the management of maternal or fetal diseases. In addition to preclinical ADME tests, studies are suggested to delineate pharmacokinetics within the maternal-fetal-placental unit.

Effects on uterine blood flow should be assessed because of the importance of this parameter for considerations of safety. A current method which permits this assessment uses chronically catheterized sheep. Studies of drugs designed for direct administration to the fetus should be conducted in animals with the

development of distribution and dose-response interrelationships. For clinical studies, evaluation should be carried out in those instances in which maternal or fetal disease warrants use of the drug. The first patients who undergo this phase III type of study should have careful evaluation of fetal heart rate via continuous electronic monitoring. Other physiologic parameters of the fetus should be followed during the period of drug administration insofar as technology permits. These pregnancies should be carefully followed, and the outcome should be meticulously ascertained - irrespective of whether the drug is administered for the duration of pregnancy or not. The infant should be carefully followed after birth until psychologic and physiologic development can be satisfactorily assessed. The state of fetal well-being should be assessed throughout pregnancy after the drug has been administered, whether singly or on multiple occasions, by measurement of urinary estriol excretion. Intrauterine growth should be assessed via noninvasive techniques, such as ultrasound. Pregnancy should be monitored by whatever means are technically available, commencing with the initiation of drug administration. This will permit determination of the time at which adverse effects occur, should such events take place. Evaluation of drug disposition will be greatly aided during this stage of development if advantage can be taken of pregnancies terminated by abortion by purposefully administering the drug just prior to termination.

Evaluation of drugs to be used for the management of labor and delivery--At this stage of development, direct assessment of effects of the drug on fetal physiologic processes (heart rate, respiration, activity) are possible, as is determination of concentrations of the drug and possible biochemical alterations (pH, glucose, etc.) in the fetus via sampling of scalp blood. Infants should be intensively evaluated at birth and throughout the neonatal period, with particular attention paid to their adaptation to extra-uterine life. This includes examination of acid-base status, weight gain, feeding ability and general activity, assessment of behavior by direct observation and through the use of psychometric tests which are valid for the neonatal period, and electroencephalography (EEG). Pharmacokinetic studies regarding drug disposition, metabolism and elimination should also be undertaken in these infants because they will have received the drug transplacentally shortly before birth. Determination of biologic half-life, excretion of the drug and its metabolites (including identification of the major metabolites in urine), and assessment of pharmacodynamic effects of the drug, if present, may be important for certain agents. Since most agents used at this stage of development are analgesics or anesthetics, careful examination of the functioning of the central and automic nervous systems is indicated. By intensive and comprehensive investigation of a few infants, followed until assessment of drug effects on psychologic and physiologic development can be made with validity, a determination can be made about the advisability of continuing trials of the drug during labor and delivery.

In the pregnant human female, studies at this stage of development can be undertaken by several different approaches. Women who receive the drug for therapeutic purposes and happen to be pregnant should be noted. Despite attempts made to avoid this situation, it will occur. The utmost advantage should be taken of this situation. Infants exposed in utero in this manner should be carefully examined at birth and followed with extensive psychologic and physiologic evaluation. This will enable ascertainment of adverse effects other than those noted at delivery. Evaluation at delivery usually detects only gross anatomic malformations.

The second approach to drug evaluation during this period of intra-uterine life involves administration of the drug to the mother, usually as a single dose, when termination of pregnancy is planned. In this instance, drug distribution, localization within the fetus, and metabolism within the fetal-placental unit

can be examined. Metabolic products should be defined within the fetal-placental unit to determine whether drug biotransformation differs from that occurring in the adult. The use of radioisotopes may be permissible because of the termination of pregnancy. In cases where there has been repeated administration of a drug to treat a maternal illness, and subsequent therapeutic or elective abortion occurs, careful histopathologic study of the aborted fetus may detect adverse effects on organogenesis.

A third approach involves careful assessment of infants receiving the drug in utero because potential therapeutic benefit for the mother was sufficient to warrant the unknown risk involved in drug administration to the fetus. Such infants should be examined meticulously at birth and followed carefully thereafter until such time as satisfactory evaluation of effects on psychologic and physiologic development can be made. The duration of this follow-up will depend on the availability and sensitivity of testing devices, the nature of the drug and its known pharmacologic, toxic and teratologic effects.

3. Special Problems

In the preceding paragraphs it has been implied that drugs will be administered mainly for therapeutic benefit of the mother. The same considerations which apply to the design and execution of clinical trials during phase II are applicable, including controls, randomization, etc. Pregnancy per se should not preclude women from participating in Phase III studies when potential therapeutic benefit of a new agent may be obtained. Special attention must also be given to the effects which pregnancy itself may exert on drug action during the randomization of phase III clinical trials.

Agents will be developed solely for the benefit of the fetus. Determination of efficacy and safety will be difficult, but objectivity demands careful assessment of such benefit in controlled trials following drug disposition studies in pregnant animals (including primates). The considerations of safety outlined for intra-uterine development are applicable when drugs are administered for the benefit of the fetus. Dosage may have to be altered considerably when the drug is administered directly to the fetus via either amniotomy or intraperitoneally. The diagnosis must be firmly established prior to administration of drugs for the treatment of fetal disease. In addition, potential benefit from the drug will have to be sufficient to warrant the risks of administration directly to the fetus.

B. NEONATAL (BIRTH TO ONE MONTH)

1. General

Newborn infants have been shown repeatedly to be much more sensitive than adults to various pharmacologic agents. This has been most often the result of differences in pharmacokinetic processes. A number of other basic considerations, including receptor sensitivity, may also account for this phenomenon. The few available data show some of the pharmacokinetic differences peculiar to neonates. They include differences in general metabolism, inequities caused by dissociation of gestational from maturational ages, a larger body surface to body weight ratio, variation of protein concentration and drug-protein binding affinity, the presence of fetal hemoglobin, immature renal tubular function, and changes in pharmacodynamic response. Small infants are most susceptible to changes of ambient temperature, and the subsequent decrease in body temperature may have notable effects on the rates of drug metabolism and excretion. Moreover, the major variations of fat and water content in the newborn and between individual neonates may result in differences in distribution and subsequent kinetics.

2. Safety

a. General Considerations of Safety: The alterations in absorption, distribution, metabolism, and excretion in the neonate may lead to accumulation of the drug with resultant toxicity. Modification of dosage may avoid this type of adverse effect. The unique physiologic state of the neonate (particularly during illness) and the wide ranges of such pharmacokinetic determinants as pH, blood gases, electrolytes, protein concentrations, and temperature present additional possibilities which may result in toxic manifestations. The very rapidity of change of such determinants makes it necessary to provide assay methods of minimal sample size.

b. Specific Toxicities

(i) Central Nervous System Effects: Evidence exists for the enhanced penetration into the brain of many drugs. The cardiovascular, respiratory, and thermo-regulatory mechanisms are extremely sensitive to depressive effects in the neonate. In addition, neuronal maturation, cell migration, dendritic arborization, and cell differentiation are occurring at this age and may be affected by drugs and/or their metabolites.

(ii) Cardiovascular

Cardiogenic effects - Drugs may affect cardiac contractility, rate, and rhythm, thereby causing severe or possibly fatal adverse drug reactions. This has been a particular problem with local anesthetic agents used during delivery. The neonate may also display delayed CNS depression or the induction of seizures and unexpected excitation resulting from the administration of some agents; he may also become addicted or dependent.

Circulatory adjustment occurring during the change from the intra-uterine to the extra-uterine environment may be hampered by the presence of certain drugs. In particular, closure of the ductus arteriosus may be impaired if respiratory depression results in hypoxemia and acidosis.

(iii) Metabolic Derangements: Changes in serum glucose, calcium, pH, sodium, potassium, etc. may be the result of drug-induced alterations in the infant's metabolic processes or may influence drug evaluation. Metabolic data obtained during the care of the sick newborn infant may provide valuable information in assessing safety and efficacy.

(iv) Changes in Bilirubin Kinetics: Prior to administration of any drug to the neonate, it is mandatory to study the drug in its final dosage form and, if possible, its metabolites and protein bilirubin binding. When appropriate, effects of the drug on conjugation, uptake, excretion, and enterohepatic circulation of bilirubin should be performed.

(v) Dermatotoxicity and Persorption: The topical application of pharmacologic agents to the neonate must be approached with an awareness of two peculiarities of this age group. First, the skin is more susceptible to dermatotoxicity expressed as photosensitivity and various forms of rash, including bullous eruptions. Second, the thin or absent statum corneum allows increased persorption, leading to systemic concentrations which may exert a toxic effect on other

organs (e.g., hexachlorophene and brain damage). In addition, systemic reactions (e.g., cyclopentolate with atropine-like toxicity) may result from increased drug absorption through mucous membranes.

(vi) Gastrointestinal: Evaluation of the effects of a drug should include consideration of such adverse effects as the inhibition of gastrointestinal motility, change of flora, vomiting, or a malabsorption-type syndrome caused by direct irritation, as well as effects on absorption of nutrients.

(vii) Hematologic: Methemoglobinemia, thrombocytopenia, and hemolysis (especially in G-6-PD-deficient neonates) may be induced in the neonate necessitating investigation of this potential in the evaluation of new agents.

c. Drugs in Breast Milk: Most, if not all, drugs administered to the mother are excreted in the breast milk. Concentrations of the drug and/or of its metabolites should be determined with due regard for the individual variations of lactation volume itself. The mere presence of the agent in the breast milk does not necessarily indicate any effect on the neonate, deleterious or otherwise, and should not in itself mitigate against approval for use in lactating women. Various factors such as concentration, the total dose delivered, the absorption by the infant, etc. must be considered in evaluating potential effects mediated through breast feeding.

d. Delayed Effects: Consideration of long-term postmarketing studies on cognitive, behavioral and physical growth depends upon the nature of the drug.

3. Efficacy

Survival rates from severe illnesses such as neonatal sepsis, idiopathic respiratory distress syndrome, erythroblastosis fetalis and hemolytic disease of the newborn, and necrotizing enterocolitis may be the only measures of efficacy available.

4. Special Problems

Some major obstacles to be overcome in establishing efficacy and safety in this age group are:

a. The Influence of Maternal Disease: The variations in physiologic states of the neonate, secondary to the pathophysiologic conditions of the mother (e.g., infants of diabetic mothers) may (1) negate the random assignment of infants to controlled, matched study populations, and (2) alter the pharmacologic response of the infant to an administered agent.

b. The Influence of Infant Disease: The wide variability within each disease state and the relatively small population of affected individuals in any single institution, together with the marked influences of the host subject in terms of gestational and maturational ages, etc., present limitations in study design, random assignment, statistical analysis, etc.

5. Ethics

The neonate presents a number of unique ethical problems. Among these are:

a. The possibility of unusual toxicity and the extreme difficulty in identification of such a problem. The late appearance, the inability of the subject to exhibit common early signs of toxicity, and the inability to verbalize symptomatic complaints all contribute to the dilemma.

b. The higher risk potential inherent in this population dictates the most substantial evidence of benefit to be derived from the use of a new drug.

C. INFANT/TODDLER (1 MONTHS TO 2 YEARS)

1. General

This period is characterized by notable increments in physical growth and rapid maturation of all organ systems with associated functional change. Noteworthy in these regards are the central nervous system and the immune system. Of direct relevance to the effect of a drug on infants in the early months of this age group are alterations in protein binding and drug metabolism.

2. Safety

a. Immediate Drug Toxicity

(i) Difficulty in detecting toxicity by clinical assessment: Toxicity may or may not be apparent in infants, especially in the early months of this age group. This may be particularly true for central nervous system toxicity. Therefore, blood levels of pharmacologic agents should be monitored and cautiously interpreted because therapeutic blood levels for older children and adults may not be safe for infants.

(ii) Gastrointestinal tract: Acute and chronic gastroenteritis is frequently encountered in this age group. Certain drugs are more likely to cause diarrhea in infants than in older children. Gastroenteritis will affect drug absorption and may complicate interpretation of efficacy and toxicity. Dehydration with resultant hypovolemia, a frequent consequence of gastroenteritis in infants, may affect drug distribution and serum concentrations.

(iii) Central nervous system: Drugs may affect myelinization and brain differentiation, which are actively occurring in children of this age group. Such effects may not be limited to drugs which localize in the central nervous system or which exhibit a predominant effect on the brain.

b. Delayed Reactions

(i) General: Toxicity is difficult to assess in this age group by clinical observations alone. Furthermore, it may not be possible to distinguish adverse effects following any single dose in a repeated series of drug administrations because of delayed reactions. Although this problem also applies to older age groups, it is particularly pertinent to infants because of their relatively immature organ systems and their limited ability to communicate.

(ii) Hypersensitivity: In this stage of initial exposure to foreign protein (e.g., foods and inhaled particulate protein), drugs may predispose to hypersensitivity through such diverse mechanisms as inhibition of secretory antibody production or induction of partial blockade of beta adrenergic receptors.

(iii) Physical growth: Physical growth may be affected by various classes of drugs such as adrenocorticosteroids and tetracycline antibiotics. Consideration of long-term postmarketing studies on cognitive, behavioral and physical growth depends upon the nature of the drug.

3. Efficacy

Although easier than for the neonatal age group, evaluation of efficacy is far more difficult than in adults. Infants cannot cooperate in a number of commonly used tests of pharmacologic action; therefore, indirect parameters (e.g., length of illness, length of hospital stay, frequency of complications and subsequent disability), and certain laboratory tests will, of necessity, be used to determine efficacy.

4. Special Problems

a. Deficiency States: The presence of iron-deficiency anemia and diminished concentrations of certain serum proteins is more likely to occur in this age group than in any other age group. Such deficiencies may alter drug kinetics.

b. Breast-feeding: The possibility of interaction from chemicals, hormones, and drugs in breast milk should be considered when suckling infants participate in drug evaluation.

5. Ethics

Before evaluating new drugs in infants, substantial evidence of benefit or superiority over accepted agents should be demonstrated in older children and adults because infants may have a higher risk potential. Included among these increased risks are those pertaining to physical growth and neurological and intellectual development.

6. Other - Research Needs

Certain research needs can be identified as relevant to the study of new drugs for this age group. (a) Relatively noninvasive techniques for determining blood levels (e.g., salivary drug concentration) should be sought; (b) noninvasive techniques for establishing efficacy of a drug should be developed; (c) much additional information is needed on the effect of drugs on the development of the immune response (both humoral and cellular components).

D. CHILDHOOD (2 YEARS TO ONSET OF ADOLESCENCE - 12 YEARS)

1. General

This age group is characterized by slower growth and the highest incidence of infectious diseases. Increasing motor and social independence results in exposure to environmental hazards which lead to various accidents such as poisoning, burns, drowning, and physical trauma. Cognitive processes involved in school performance and school attendance - vital to intellectual and psychosocial development - are being rapidly acquired. At the end of this age period, rapid bone growth and epiphyseal maturation occur secondary to changes in endocrine activity. Accordingly, pharmacokinetics may differ from the infant and adolescent age groups, depending on the characteristics of the drug and the child's age within the broad age range of this period.

2. Safety

a. General: Safety considerations in general differ little from those in Section I. A specific need at this age, when accidental poisoning is common, is information dealing with acute toxicity and treatment of drug poisoning.

b. Specific Toxicities

(i) Immediate drug toxicity: A disease for which a drug is given may enhance its toxic potential. Thus, interaction with disease states which would apply particularly to drugs used at this age should be studied, e.g., antibiotics, bronchodilators, antihistamines, and anticonvulsants. An example would be the altered toxicity of ampicillin when employed in infectious mononucleosis or increased toxicity of isoproterenol (ventricular tachycardia) when the patient has hypoxemia and acidosis.

Hypersensitivity manifested by anaphylactoid and anaphylactic reactions are more likely to occur at this age and in adolescents than in younger children because of longer periods for sensitization and greater exposure to antibiotics and similar substances to which antibodies may be induced.

(ii) Delayed Reactions

Hypersensitivity manifested by serum sickness or drug fever--This may be seen with a variety of agents ranging from antibiotics to anticonvulsants and is common in this age group and in adolescents.

Drugs interfering with school performance and other childhood activities--These may include, but are not limited to, side effects which interfere with attention span (e.g., drowsiness) or reduce perception (e.g., tinnitus and decreased hearing).

Drug-nutritional interactions--The prolonged use of a drug in a child may affect his nutritional requirements. Recent observations on the rachitic effect of long-term administration of diphenylhydantoin illustrate this concern.

(iii) Late Onset Reactions

Chronic administration of a variety of agents may affect linear growth and/or weight gain.

Selective growth changes include advancement or retardation of puberty or of menarche.

3. Efficacy

Evaluation of efficacy based on objective criteria is possible in the school-aged child who is able to cooperate. Objective measurements should be stressed in study design. School performance and school attendance provide additional parameters which may be extremely useful in determining efficacy. Even though the rate of physical growth has slowed in this age group, changes in growth rate may provide additional evidence of efficacy, especially in those diseases which depress linear growth or interfere with normal weight gain. Assessment of osseous development (e.g., bone age) is one parameter of growth that may be useful where indicated. The efficacy of agents in preventing or altering morbidity from infectious diseases may be best studied in this age

group when the incidence of viral and bacterial infections is high.

4. Special Problems

Accidental poisoning and overdosage are of prime consideration at this age. The manifestations of acute poisoning with the drug and its metabolites can be studied in juvenile animals. Information concerning specific antidotes and therapy of overdosage (e.g., peritoneal dialysis) should be included in the protocol and ultimately in the package insert.

5. Ethics

Special ethical consideration in this age group involves school absenteeism for studies as well as the psychological effects of such studies on the child. These should be discussed with parents before informed consent is obtained. Older children may be able to participate in the consent process.

E. ADOLESCENT (ONSET OF ADOLESCENCE TO ADULT LIFE - 12 TO 18 YEARS)

1. General

Adolescence may be defined as the transition period in which the child undergoes changes in physical, sexual, and psychosocial development transforming her/him into an adult. During this time period, the child's body is rapidly changing in form, undergoing final rapid growth to mature stature and the development of secondary sexual characteristics. Coupled to the dramatic changes in body form, the adolescent develops a new perception of her (him)self as an individual in relation to her/his niche in the family and in the general fabric of society.

Changes in physiology may produce alteration in the absorption, distribution, metabolism, and excretion of drugs as well as in receptor respose. The development of puberty and the known effects of sex hormones on drug metabolism warrant consideration in drug evaluation in the adolescent.

2. Safety

a. General Considerations of Safety

The major concerns relating to drugs given to an adolescent involve:

(i) the potential for abuse;

(ii) the possibility that the agent may alter the final stages of physical and endocrine development completing the growth cycle to maturity.

In addition, in this age group, medication may not be taken as prescribed. The adolescent frequently omits doses of medication, takes it at erratic intervals, and may take more than prescribed. Safety considerations should be addressed not only to the therapeutic dosage, but also to the consequences of suboptimal dosage and overdosage.

b. Effect of the Age Group on Safety Considerations

(i) Immediate Adverse Effects

Drug misuse includes that of accidental or intentional overdosage or underdosage and that of inappropriate use. The adolescent may fail to take the medication as frequently as prescribed, or he may

employ it in larger doses than prescribed or for inappropriate reasons. The effects of such practices on the disease process and adverse effects will have to be anticipated.

Hypersensitivity reactions include anaphylaxis, serum sickness, and contact dermatitis. Although not unique for the age group, these reactions may occur as a result of self-medication or inappropriate routes of administration of medication.

(ii) Delayed Reaction

Dependency and habituation are among the major delayed reactions.

(iii) Late Adverse Effects

Psychosocial and behavioral alterations may occur as a late, even unexpected, action of a drug and should be considered in drug evaluation. These may occur either as a direct effect or as an exaggeration of an underlying problem.

Other--Growth changes, advancement or delay of puberty and of menarche, and effect on fertility may consitute other delayed drug reactions in this age group. Consideration of long-term post-marketing studies of possible drug effects in these areas depends upon the nature of the drug.

Pregnancy test on female participants--Because of the presence of unknown or hidden early pregnancy, adolescent girls should have pregnancy tests before entering any drug trials.

3. Efficacy

The same objective measurements used in adult patients to define efficacy should be used.

4. Special Problems

a. General: The plasticity of evolving form and functions in the adolescent produces unique therapeutic problems for this age group which can be grouped into three major categories.

(i) Drugs used to alter physical growth and sexual development. Drugs given to regulate growth or secondary sexual manifestations are unique to the adolescent. Many pharmacologic agents are employed in an attempt to make the subject "normal" or "superior" regarding growth, muscular development, or sexual development. Pressures to use drugs are generated by the adolescent's peer group. An adolescent who is too tall or too short, too obese or too thin, or not athletic enough is made the object of derision by his or her peers. Synthetic androgens are often used under these circumstances. Their effects on hepatic function (and metabolism of other drugs) and hepatic carcinogenesis should be taken into consideration.
The problems of potentially tall girls and of irregular menses may both be treated with synthetic sex hormones. The long-term effects of these practices must be studied with regard to fertility and carcinogenesis. The latter is highlighted by the development of uterine carcinoma in patients with Turner's syndrome after stilbestrol treatment.

Conditions affecting both males and females are obesity and sexual precocity. Growth and fertility could be affected by agents used in their treatment. For example, medroxyprogesterone - used in treatment of sexual precocity - has been shown to suppress the pituitary-adrenal axis, cause Cushingoid features, and produce "sticky-chromosomes" in the male gonad. These examples of adverse effects warrant consideration when new drugs of this class are evaluated.

(ii) Drugs used to regulate mood and behavior. The adolescent is prone to psychosocial disturbances; the ambivalence created by his/her striving for self-identity and his/her dependent needs coupled with rapid changes in physiology and body form create a milieu of stress. Bizarre and unusual behavior may result when family interrelationships are strained or if school and peer interactions break down. Depression, anxiety, and acting out are common psychological symptoms which the physician is requested to control with drugs. There the problem of evaluating efficacy may be confounded by concurrent psychotherapy; this must be considered when adolescents are enrolled in a psychoactive drug study.

Effects on school performance, social behavior, and operation of vehicles should be kept in mind.

(iii) Drugs used for cosmetic purposes. Awakening interest in the opposite sex is characteristic of the adolescent. The adolescents' self-image in this context is related to their physical attractiveness. Minor skin blemishes may result in an inordinate expenditure of effort, time, and money to correct anything which may be considered a defect. At the same time, physiological changes make them susceptible to acne, seborr'ea, and hirsutism. They seek and use a variety of medications, both on prescription and over-the-counter, to contend with these problems. Antibiotics, hormones, and vitamins may be prescribed for systemic use or topical application. Other medications (such as keratolytics, drying agents, and ointment powders to cover blemishes) are limited to external use.

For topically applied drugs, the problem of skin sensitization is superimposed on those of potential abuse and overdosage common to other classes of drugs.

5. Ethics

a. Informed consent should be obtained from the subject as a responsible individual, as well as from her/his parents.

b. The effects of drugs, even in the young adolescent, must include the possibility that females are pregnant and males may be fertile.

c. The possibility that the drug may have an effect on ova or spermatozoa must be considered.

6. Other - Compliance

Patients may fail to take the medication under study according to their protocol. This is particularly true of adolescent patients who are not yet mature enough to realize the need to take even the most important medications (i.e., insulin in juvenile-onset diabetes). Therefore, to evaluate drugs in this

age group, methods to evaluate compliance will have to be devised and used.

III. SUMMARY OF REQUIRED STUDIES

The following summary is intended to list those studies which are felt to be required in all (or almost all) drugs to be approved for use in pregnant women, infants, and children. There will be exceptions. The recommendations are divided into two groups: animal studies and studies in pregnant women, infants, and children.

A. STUDIES IN ANIMALS

1. Chronic toxicity studies. This is the usual long-term multidose administration to two species, usually the rat and beagle dog. These studies should include effects on growth and skeletal maturation (bone age).

2. Appropriate methods for determining bioavailability using nonradiation-emitting techniques are to be developed. Initially "hot" methods for animal studies may serve as a prototype for the development of appropriate "cold" methods, but efforts should be directed to developing a sensitive "cold" method. The method(s) should be sensitive enough to measure with small sample size levels in serum expected to be in the therapeutic range. The method(s) should also differentiate the drug from its major metabolites. If the latter are pharmacologically active, additional techniques for these measurements are needed.

3. The pKa and lipid: water ratio of the chemical moiety used in the product should be determined.

4. Studies of absorption, distribution, metabolism, and excretion. These should account for a major percent of the administered dose and lead to formulation of a pattern of metabolism and disposition during both acute and chronic administration. Major metabolites should be identified. Unusual disposition -particularly in growing bone, teeth, or endocrine organs - which might be associated with adverse effects in the pediatric population should be sought.

5. The standard "3-phase" reproduction study.

B. STUDIES IN PREGNANT WOMEN, INFANTS, AND CHILDREN

The following factors are to be determined in each age group for which the drug will be approved. The usual sequence of testing should first involve teen-agers then successively younger children. Exceptions will occur when diseases are peculiar to one age group. The neonate must be approached with great care, since even studies in young children may not yield a reliable estimate of toxicity for the neonate. For studies of the fetus, infants treated as an inadvertent recipient by administration to the mother of a drug for a serious medical problem may be the first studies involving the fetus. Throughout the recommended studies that follow, there apparently are no important sex differences before puberty; thus, data obtained from both males and females may be pooled. This is a reasonable but still untested postulate, however.

1. Blood levels found with the range of doses adopted from studies in adults. If such studies have determined the therapeutic range, the dose required in infants and children to achieve this range must be an early priority.

2. Studies of absorption, distribution, metabolism, and excretion. The goals of such studies should include localization in tissues, rapidity of excretion, and time of peak onset.

a. Absorption. The percent of a single and/or multiple dose that is absorbed should be determined.

b. Distribution. Binding to plasma proteins at therapeutic blood levels should be determined. Studies of displacement of bilirubin from serum albumin are critical if the drug is to be used in neonates or late in pregnancy. If such displacement is found, additional studies with drugs which may be concurrently administered and the effect of pH, free fatty acids, etc., on the drug albumin-bilirubin complex are mandatory.

c. Metabolism. Determination of the major biotransformation products, including a search for unique or unusual metabolites, may be coupled with studies of blood levels (No. 1). If significant age-related changes are found in metabolism, then a comparative profile of quantitative changes occurring with age may be necessary.

d. Excretion. The fate of the drug, expressed either as percentage of the multiple daily dose or as single dose with an appropriate time scale as determined from the decline in serum levels or other monitor of excretion, should be ascertaind. Such studies should account for a major protion of the administered dose in most instances.

3. Bioavailability. If the dose form to be used in children is significantly different than that for adults, it must be considered as a new drug, and absorption and excretion studies should first be performed in adults. In any event, the dose form or forms used for pediatric patients must be used for studies of absorption in children. This stipulation will cover the potential problem of toxicity or influences of the vehicle or other components of the formulation.

4. Because of the multiple unique aspects of the neonate, a neonatologist should be part of the team which evaluates the influence of a new agent to which a fetus or a neonate has been exposed. Study must be made of possible interferences by the drug with metabolic reactions unique or of particular importance to neonates, such as the handling of bilirubin, glucose homeostaasis, acid-base balance, oxygen-carrying capacity, development of pulmonary surfactant, etc.

5. Depending upon the drug, consideration should be given to establishing a program for long-term follow-up of the offspring of women receiving the drug during pregnancy. Such studies need to evaluate both possible intra-uterine death and malformations. Since many malformations are not detected at birth, a program of follow-up should insure evaluation at least at 1 year of age. Malformations should include functional as well as anatomic abnormalities. Even longer follow-up is desirable, particularly for drugs which might be anticipated to have an adverse effect on neurologic development. However, the difficulties of such long-term studies are recognized and some compromise must be made. Depending upon the drug, similar but perhaps less intensive and extensive follow-up may be needed for children receiving the new drug during postnatal and later developmental stages.

6. For drugs which may be used chronically, the effects on weight gain, statural growth and skeletal maturation (including, perhaps, in some cases, serial bone age films), and sexual maturation should be assessed. The effects of chronic administration on behavior and learning are important areas, yet ones in which no exact requirements for studies can be delineated. The determination of effects on behavior and learning may be part of the evaluation of efficacy of psychoactive compounds; thus, indirectly, some data on safety will be

obtained. However, in addition to specific benefical effects which will be observed, other areas demanding consideration are:

a. classroom attentiveness and performance,
b. grades, comments of teachers, etc.
c. unusual or bizarre behavior,
d. somnolence, depression, withdrawal,
e. reports of trained observers, parents, teachers,
f. formal testing procedures.

In general, the longer the drug is to be administered the more important long-term follow-up becomes.

7. Studies of hematologic, hepatic, and renal damage from acute and chronic administration are needed because these organs are most readily affected by drugs, even if no toxicity has been demonstrated in adults. Such studies must be done with acute and chronic dosing.

8. Depending on the drug, specialized studies such as ECG, EEG, hearing, vision, etc. may be required. Certain clues can be taken from studies in adults and from the pharmacologic and chemical nature of the drug in determining the number and extent of such studies.

9. Before investigations are begun, provision must be made available for management and treatment of accidental or intentional overdosage and for severe toxic reactions to the drug.

10. Data must be obtained on the influence of the drug on fetal growth and differentiation for drugs which will be approved for pregnant women. Apgar scores, performance in the nursery, etc., are necessary parts of such studies. When appropriate, studies of addiction of the neonate and presence of withdrawal signs or symptoms must be performed or be in progress.

11. Concentrations of the drug and/or its metabolites in breast milk and effects on the nursing infant should be determined for drugs to be used in lactating women.

All recommendations made throughout these guidelines - and particularly in this summary section - must be viewed from the standpoint of flexibility, and appropriate modifications should be made for the individual drug, its indications for use, and the age of the patient for which it is intended.

APPENDIX III

DIAGNOSTIC CLASSIFICATION IN CHILDHOOD AND PEDIATRIC PHARMACOLOGY*

The focus of this appendix is on the proposed classifications for Disorders of Childhood and Adolescence for the forthcoming revision of the APA Diagnostic and Statistical Manual. Some introductory comments are, however, appropriate.

QUALITIES OF AN IDEAL DIAGNOSIS

The minimum information a diagnosis can communicate is the phenomenology of the disorder or the constellation of clinical symptoms grouped under the common label. Doing so provides a consensus which enhances the communicative value of diagnostic terms. This grouping process can be arbitrary simply as a starting point, or based on careful, objective clinical evaluation from a variety of statistical techniques such as factor, cluster, or multiple regression analysis, which estimate the relationships of observed clinical phenomena mathematically (Klein & Davis, 1969). However, if all a diagnosis can do is to indicate a clinical picture, and no more, it has limited value indeed.

In the ideal, a diagnosis has several characteristics. It should provide a good estimate of the natural history of the disorder (outcome or prognosis without treatment); its outcome given current treatment, its etiologies; the pathophysiology of the disorder if there is a specific biological cause, and, if there is a psychosocial cause, the psychological mechanisms underlying the disorder. These prognostic and etiological data are those necessary in the end to validate the syndrome and show that it is more than an arbitrary concatenation of signs and symptoms. When all these factors are known, the ground is laid out not only for curing the disorders, but, better yet, for preventing them. Of course, the process of discovery need not always proceed in this orderly fashion and the establishment of specific treatments may help to define clinical syndromes (Klein, 1963, 1973) such as may well prove to be the case with depression or hyperkinetic disorder (Wender, 1971).

Unfortunately, very few psychiatric disorders of children have been investigated sufficiently so that it may be stated with confidence that they have the associated etiological, prognostic, therapeutic and preventive validating factors discussed above. Though this uncertainty is unfortunate, it should act to stimulate systematic research in diagnosis in child psychiatry rather than lead to a defeatist attitude. Interest in, and attention to diagnosis is especially characteristic of psychopharmacology, since a particular drug might be indicated in a specific disorder.

*By R. Gittelman-Klein, Ph.D. and J. Rapoport, M.D., adapted in part from "Diagnostic Classifications and Psychopharmacological Indications" by R. Gittelman-Klein, Ph.D., R. S. Spitzer, M.D. and Dennis Cantwell, M.D. In J. Werry (Ed.) Pediatric Psychopharmacology. Bruner/Mazel, New York, in press, 1978. The authors gratefully acknowledge Dr. John Werry's editorial contributions.

IMPROVING DIAGNOSIS

A close relationship between drug treatment and disgnosis is not common in pediatric psychopharmacology. For example, the antipsychotics such as the phenothiazines are not effective as specific antipsychotics in children, but are used symptomatically to reduce motor activity in overactive children, regardless of any diagnostic considerations.

The degree to which the desideratum of the "right drug for the right patient" can be met is, in part, a function of how reliably we can measure the child's behavorial signs and symptoms, on which, at the moment, diagnosis in child psychiatry largely rests. There is certainly room for significant improvement in current diagnostic practice, but the utility of diagnoses also rests on the ability of the symptoms to reflect discrete clinical categories meaningful for pharmaco therapy or other interventions. Is there indeed still a better chance of finding the right drug after having identified the right diagnosis through an improved taxonomy? An obvious example of the value of a good taxonomy is the use of lithium in manic-depressive disorders. The syndrome was identified long before the use of the drug and thus facilitated the discovery of lithium treatment in those disorders. It is hoped that the improvements in the current diagnostic system for children (which most are agreed is unsatisfactory) may set the stage for the discovery of relationships between specific disorders and specific treatments (not necessarily all pharmacological).

Arguments have been advanced in favor of another approach to diagnostic classification whereby pattern of drug response would be used as the basis for identifying homogeneous patient groups (Klein, 1973; Wender, 1971). However, this strategy is now at variance with the phenomenological approach in diagnosis, and the two should complement each other.

ALTERNATIVE CONCEPTS IN DIAGNOSIS

Medical diagnosis, of which present psychiatric classifications are part, is basically qualitative and dichotomous-the patient either has the condition or does not and the disorders are qualitatively different from each other. In contrast, behavorial scientists often espouse a dimensional approach to diagnosis in which behavior is believed to consist of a number of dimensions rather like height and weight, along which any individual's behavorial or personality profile may be plotted in N dimensional space. Abnormality then is simply an extreme position on one or more dimensions with the "syndrome" being defined by the profile rather than a simple YES/NO categorization. This method has been used extensively with personality tests like the MMPI and in children's behavior rating scales (Quay, in press). It is of interest that, when used as a classification device, this is usually achieved by truncating extreme scores into "profiles" or syndromes which have all the features and assumptions of the medical model. For example, Quay's "conduct poblem" child seems much the same as DSM II's Unsocialized Aggressive Reaction.

Ordinarily, and probably more properly, these dimensions have been used in pediatric psychopharmacology as predictors or measures of drug effect rather than as diagnostic categories. The new APA classification (DSM III) has some dimensional features but the heart is the medical model.

CHANGES IN THE APA DIAGNOSTIC SYSTEM (DSM III) COMPARED WITH DSM II

Historically, pediatric psychiatric diagnosis has received little systematic attention apart from one previous attempt to provide a comprehensive diagnostic schema for children which, perhaps because of its lack of official origin, was not widely accepted (GAP, 1966). Recently, however, official bodies have given intensive consideration to children's diagnostic systems and have advanced relatively elaborate descriptive systems for implementation shortly.

The third edition of the American Psychiatric Association Diagnostic and Statistical Manual (DSM III) was initiated with the goal of preparing a classification system based on current psychiatric knowledge. Substantial changes were made both in diagnostic approach as well as in specification of individual clinical entities.

Table 1 presents a comparison for the convenience of the reader between DSM II and DSM III for Disorders Usually Arising in Childhood or Adolescence. These are disorders which usually originate during childhood and, which, typically, are not known to have adult onsets. On the other hand, many disorders arise across a wide age span, from early childhood through adulthood; for example, obsessive-compulsive disorders may begin in childhood, but are also known to occur de novo in adulthood. Such disorders are not listed in the section of DSM III specific to childhood. Therefore, if a child presents with difficulties which coincide with those stipulated for any disorder outside the section for childhood disorders, the appropriate adult diagnosis is to be applied. Consequently, it would be erroneous to assume that the list of conditions enumerated under Disorders Usually Arising in Childhood or Adolescence represents the universe of diagnoses applicable to children. Children may receive diagnoses from the other major diagnostic rubrics included in DSM III: Organic Mental Disorders, Drug-use Disorders, Schizophrenic Disorders, Paranoid Disorders, Psychoses Not Elsewhere Classified, Affective Disorders, Anxiety Disorders, Factitious Disorders, Somatoform Disorders, Dissociative Disorders, Personality Disorders, Psychosexual Disorders, Reactive Disorders Not Elsewhere Classified, and Disorders of Impulsive Control Not Elsewhere Classified. However, DSM III specifies that a child should receive an adult diagnosis only when the categories for children do not provide an appropriate coding.

DSM III departs from the previous diagnostic system (DSM II) in general conceptual ways; in addition; DSM III differs in ways which are specifically pertinent to children's psychiatric disorders. For the sake of clarity, a discussion of the overall general discrepancies precedes that of the children's nomenclature.

GENERAL DIFFERENCES BETWEEN DSM II AND DSM III

The changes described below are not comprehensive in scope; only the major points which affect the diagnosis of children's psychiatric disorders are summarized. Many other discrepancies exist, such as those among Organic Mental Disorders, but their relevance to the pediatric diagnostician is remote and therefore they do not require attention here.

Major Categories

Certain classificatory umbrellas have been abandoned in DSM III and consequently affect the way in which mental disorders for children are organized. The classical terms Psychosis and Neurosis are no longer grouping concepts for mental disorders.

1. Psychosis.

   The term psychosis connotes a multitude of clinical phenomena and is therefore confusing. It does not represent a homogeneous group of conditions, but is a particular aspect of mental dysfunction which may occur across many types of disorders; for example, in organic mental disorders, it often refers to changes in intellectual functions such as memory and orientation; in schizophrenia it may connote the presence of abnormal ideas or preceptions, delusions and/or hallucinations; in affective disorders the concept of psychosis may be applied to patients with delusional ideation, or alternately to those severely dysfunctional in mood, but with no delusions; finally, it can be applied to the individuals who, under the influence of psychotomimetic or other drugs, undergo marked changes in their experience of reality. Therefore, the notion of psychosis as a class of disorders is untenable. Further, in children, the term psychosis poses additional definitional problems. Adult psychotic individuals are usually not mentally retarded; they are thought to have achieved a certain level of adjustment (with varying degrees of

adequacy) and at some time to have undergone marked changes in personality. This picture does not apply to children with autism which is the commonest psychosis of childhood (Werry, in press), since, typically, they have not experienced a period of social normalcy interrupted by psychosis, but are developmentally deviant from infancy on, and are frequently intellectually retarded. In addition, and of some importance, the therapeutic connotations which the term psychosis often carries in adults is misleading. The neuroleptics have come to be known as antipsychotic agents. Given their documented efficacy in adult schizophrenia, this practice may not be wholly unjustified. However, their action in so-called psychotic children cannot be inferred from adult "psychotics" since the medications do not normalize the behavior of the children in the same fashion as that of adults. Finally, the family histories and manifest clinical symptoms of most psychotic children and adults are markedly different (Werry, in press). Therefore, there seems to be little point to using the same diagnostic term in children and adults with different respectige phenomenologies, different prognosis, different treatment indications and probably different etiologies.

2. Neurosis.

   Neurosis as classificatory concept was abandoned for different reasons. Unlike psychosis, it carries etiological inferences-the assumption being that the overt neurotic symptoms represent superficial manifestations of unconscious, repressed intrapsychic conflicts, the exact nature of these conflicts varying with the psychopathological theory of the diagnostician's predilection. Since it was agreed upon that etiological speculations would not be part of the new objective descriptive diagnostic schema, retaining the concept of neurosis was unjustified.

3. Hysteria.

   The term Hysteria, a subset of the neuroses in prior classifications, has also been removed from the nomenclature in DSM III. As neurosis, it too implies causal factors which remain unproved and has acquired confusing and often pejorative connotations. In its place, more descriptive labels and more specific disorders formerly grouped within "hysteria" have been included under the rubrics of Factitious Disorders, Somatoform Disorders, and Dissociative Disorders, which it is hoped are novel enough to have some chance of exact meaning.

Diagnostic Descriptors

The content of all of DSM III diagnostic descriptions will include a summary of the following descriptors:

1. Primary clinical features;
2. Frequently but inconsistently associated secondary symptoms;
3. Age at onset;
4. Course of the disorder;
5. Complications;
6. Predisposing factors;
7. Familial pattern
8. Prevalence
9. Sex ratio

10. Differential diagnosis; and as discussed below in more detail,

11. Diagnostic criteria for making the diagnosis. No such comprehensive attempt was made for DSM II.

Diagnostic

A critical step forward has been taken by DSM III by providing operational criteria for each disorder so that the clinician will have guidelines to apply in the diagnostic decision process. An obvious question is how can these criteria be formulated in the present absence of objective empirical data which define the limits of each disorder? There is no wholly staisfactory answer: yet, there has to be a point of beginning somewhere. Therefore, those clinicians involved with the development of DSM III formulated a set of arbitrary, but it is hoped, sensible rules based on current knowledge and experience. As a result, the objectivity of the operational criteria is very variable. In cases where considerable information regarding a condition has been accumulated, the job of formulating criteria was both easier and more rational. In contrast, where disorders have been included because of a consensus that the category exists, but no systematic studies have been conducted, the criteria are arbitrary. In practice, it will be difficult at times to be certain whether the criteria for a particular diagnosis are met by some patients. Thus, a patient might fit some, but not all the criteria of a particular diagnosis. In such instances the diagnosis should be used if it appears clinically to be the best diagnostic fit possible.

Operational criteria improve diagnostic reliability, and will render investigations of the validity and epidemiology of mental disorders more feasible.

Finally, the criteria will probably enable individuals without extensive training in psychodiagnosis to use the manual with less confusion and ambiguity, thereby making it more useful to pediatricians and other non-mental health professionals involved in the care of children. The proposed content of DSM III is being subjected to clinical trials in clinical settings before finalizing its contents. Doing so will identify the disorders which, though proposed, are not readily applicable to patients; it will identify some of the ambiguities of the descriptive content and the operational criteria; it will point to omissions by identifying patients who cannot be diagnosed by DSM III.

Therefore, there will be an opportunity to revise the manual prior to publication and any difficulties identified can be removed and need not wait a decade to be incorporated in the next manual. The present system is tentative, therefore, as the final classification will not be ready until 1979.

MULTIAXIAL CLASSIFICATION SYSTEM

The Multiaxial diagnostic approach proposed by some of those involved in ICD 9 under the aegis of WHO (Rutter et al., 1973, 1976) has been adopted for childhood disorders in DSM III, and it will be applied to adults as well as children and be an integral part of the diagnostic coding.

Advantages of Multiaxial Classifications

It should be made clear that a multiaxial classification in no way precludes the use of several clinical diagnoses, so that a child with an Attention Deficit Disorder and a Conduct Disorder would receive two codings to reflect the presenting psychiatric symptomatology. The multiaxial approach simply insures that certain specified domains of function are regularly assessed in all cases and thus should improve diagnosis in child psychiatry.

Besides the research advantages of providing a large pool of cases evaluated along similar dimensions, commonly assessed at present, but obscured in current systems, the multiaxial coding maximized diagnostic reliability. For example, it has been demonstrated (Rutter et

al, 1973) that when children presented with mixed clinical pictures such as a concatenation of severe behavior disorder, epilepsy, and mental deficiency, categorization was inconsistent since primary consideration and weight was given to different aspects of the children's conditions by different diagnosticians. Regardless of which of the three classes of dysfunctions was chosen as the clinical diagnosis (epilepsy, psychosis or mental retardation), a correct decision was actually made, but, in each instance, being uniaxial in nature was incomplete and seemingly in complete disagreement with any of the other diagnoses. However, given multiaxial ratings, no such diagnostic confusion should occur. Therefore, the system should enhance the validity of the diagnostic decisions, always assuming, of course, that the axes selected reflect attributes which are relevant either to the origin, course, or treatment response of the disorder. Even if this ambitious clinical goal is not met with the use of multiple axes, the latter will still be useful by enhancing the scope and accuracy of descriptive content communicated by the diagnosis formulated--a small advance, but a real one nonetheless. As shown in Table 1, Axis I reflects the clinical disorder, and, as noted, multiple diagnoses may be used. Axis II is for ratings of specific developmental disorders (also on Table 1). These are deficiencies in development which cannot be attributed solely to mental deficiency or gross deprivation (such as absence of schooling) and encompass such problems as specific learning disabilities, motor incoordination and delays in bowel and bladder control. Not shown here are Axis III, IV and V of the proposed classification scheme. In DSM III, Axis III provides the opportunity to note the concurrence of physical or biological disorders which are felt to be, in some way, pertinent to the clinical psychiatric condition by affecting its course, severity or management. This association need not be clearly established for the former to be recorded. For instance, the presence of epilepsy, diabetes, or asthma would be coded even when no obvious link existed between them and the psychiatric syndromes, since their association with psychiatric disorders is not infrequent. However, clarly transient, acute medical disorders would not be noted except under extraordinary circumstances. On Axis IV of DSM III, the diagnostician notes whether a significant psychosocial stress appears to have contributed to the clinical disorder and if so, the severity of the stressor, from minimal to catastrophic. In contrast to Axis III, an etiological relationship between the presence or severity of the psychiatric disorder and the stress is inferred.

Axis V of DSM III indicates the highest level of the patient's psychosocial functioning during the year preceding the evaluation. This aspect of adjustment is felt to be often important in planning treatment. The behavior rated on Axis V is independent of other clinical considerations such as overt symptoms, or subjective distress, and reflects excusively the patient's level of adaptive functioning, rated on a scale from an optimum or "superior" to a minimum of "grossly impaired."

Differences in Diagnostic Categories

In addition to the DSM III categories presented in Table 1, over 100 additional diagnoses not listed as specifically arising in childhood also may be used for children. These include Reactive Disorders, Sleep Disorders, Psychosexual Disorders, Schizophrenic Disorders and Somatoform Disorders which, it is anticipated, will be regularly used with children and adolescents.

It is apparent that, on the whole, the sheer number of discrete diagnostic categories is greater in DSM III than DSM II; this stems from the philosophy which guided DSM III codings --any distinct, internally consistent phenomenological symptom pattern merited its own code, so that more knowledge could be accumulated for the disorder. Diagnostic refinement was felt to be unlikely to lead to loss of information, whereas lumping discrete conditions on purely traditional grounds might obscure important clinical differences among the combined disorders.

The danger in this splitting process is that it can go too far. There is an awareness of, and no wish to revert back to, the chaos of pre-Kraepelinian European psychiatry, with a myriad of disorders all lacking in validation. Fortunately, in contrast to then, there are now investigational tools available to enable rapid and sophisticated documentation of the validity of

a category--for example: the research technology of epidemiology, drug responsivity, outcome, genetrics, and inferential statistics to say nothing of electronic data processing and a vastly increased body of psychiatrists and knowledge.

CHILDHOOD DISORDERS AND THEIR PHARMACOTHERAPY

The disorders discussed below follow the DSM III classification. The brief note regarding the pharmacological management of the respective DSM III categories is not intended to provide a comprehensive view of the field. Rather, it is mentioned only to point out existing relationships between pediatric psychopharmacology and the proposed new systems. Obviously much more work will need to be done to extend this to cover new and/or pharmacologically untested diagnostic categories and subcategories.

MENTAL RETARDATION

Diagnosis

The coding of mental retardation is considered to be a clinical psychiatric diagnosis in DSM III and therefore appears on Axis I. The DSM III diagnosis of mental retardation requires subnormal IQ (at least two standard deviations below the mean), but, in addition, a concurrent deficit in the capacity for adaptive behavior. Given this bivariate definition of mental retardation, a child with an IQ below 70 who was managing well in meeting the usual role expectations for his age, such as going to school (though in a special class), self-care, and so on, would not be considered to have a mental retardation in DSM III. These DSM III criteria were adopted so as to be consistent with those of the American Association for Mental Retardation (Grossman et al, 1973). However, the severity ratings for Mental Retardation in DSM III are strictly dependent on level of quantified IQ, and not affected by other considerations. There is some inconsistency, therefore, within the DSM III schema since the diagnostic criterion for the disorder rests on the presence of two sets of dysfunctions, but its severity only on one. This lack of internal consistency is due to the fact that ratings of adequacy of adaptive behavior are relatively subjective and therefore more unreliable than IQ measures, and including an evaluation of psychosocial adaptation in the severity codings of mental retardation would have affected the reliability of the diagnosis negatively.

PSYCHOPHARMACOLOGICAL TREATMENT

There is no specific treatment for Mental Retardation. However, children with the disorder often have behavorial problems, not infrequently severe. Symptoms of hyperactivity, aggression, destructiveness, self-damage (such as hitting, biting, banging onself) occur. The neuroleptics have been found to ameliorate these secondary clinical complications, but do not affect the primary intellectual deficit; in fact they may depress cognitive function. Stimulants, despite their facilitative role on cognitive function in the laboratory, have yet to be shown to influence academic skills and learning in general.

PERVASIVE DEVELOPMENTAL DISORDERS

Diagnosis

Infantile autism refers to severe deviance in the development of social responsiveness, occurring very early in life as originally described by Kanner (1943). The disorder has three key features: onset during the first three years of life, lack of social responsiveness, and deficits in language development.

In keeping with some of the confusion in this area (Werry, in press), the disorder of Atypical Childhood Psychosis is a less distinct grouping which will need refinement. Its clinical picture is more varied than that of Infantile Autism, the age of onset later. The term psychosis as a description for the category of Early Childhood Psychosis has been retained since, by definition, the children must display several years of relatively normal social and intellectual functioning before marked changes in object relationships are observed.

PHARMACOLOGICAL TREATMENT

The antipsychotics have been demonstrated to have some beneficial effect on the secondary signs and symptoms of this group of childhood disorders. However, the drugs do not have a normalizing or true anti-psychotic action as in adults; nor do they eliminate or reduce significantly much of the children's bizarre interests and inappropriate social interactions, but they may have dramatic beneficial effects on certain disturbing symptoms such as severe hyperactivity and mood liability. Troublesome withdrawal side effects, such as kyskinesias, have been reported and the cost benefit ratio of using high levels of neuroleptics over extended periods of time must be weighed for each child.

ATTENTION DEFICIT DISORDERS

Diagnosis

This category is for disorders often referred to as Minimal Brain Dysfunction, or Hyperkinetic Reaction of Childhood, the hallmark of these disorders being now considered (though by no means unanimously) to be marked impairment in sustained attention processes.

DSM III distinguishes between two categories, Attention Deficit Disorder with Hyperactivity, and Attention Deficit Disorder without Hyperactivity. The latter disorder refers to children with pure attention dificit disorders without behavorial problems. It is a controversial category, based on clinical reports that some children exhibit difficulty in sustaining attention and applying themselves, without any dysregulation of patterns of motor activity. The existence of the syndrome itself has not been documented. It is hoped that its inclusion in DSM III will lead to attempts at providing objective evidence for the disorder, or lack thereof. The DSM III Attention Deficit Disorder with Hyper activity is equivalent to the DSM II Hyperkinetic Reaction.

DSMIII does not provide for mixed classification of Hyperkinesis with Developmental Delay or with Conduct Disorder; in such cases, a multiple clinical diagnoses are to be used. The reason for avoiding a mixed diagnosis in DSM III is due to the difficulties in establishing a primary and secondary diagnosis in a child who presents with several patterns of dysfunction.

Pharmacological Treatment

Of all the childhood disorders, the Attention Deficit Disorder with Hyperactivity is the one for which drug treatment is best documented. The stimulants have been shown repeatedly to improve dramatically the clinical symptoms of the disorder. Some antipsychotics such as certain phenothiazines can also ameliorate motor hyperactivity, but they do not have the broad normalizing therapeutic effect of the stimulants. It is unknown whether stimulants are also useful in the treatment of children with Attention Deficit Disorder without Hyperactivity. The clinical impression is that they probably are. Tricyclic antidepressants are also useful though less so than stimulants and they may be more toxic.

STEREOTYPED MOVEMENT DISORDERS

The Gilles de la Tourette's syndrome has been split from pure motor tics (Transient Motor Tic Disorder and Chronic Motor Tic Disorder) in DSM III. It is not clear whether the distinction

is warranted. Part of the reason for distinguishing individuals who, in addition to motor tics, also have involuntary verbal outbursts, is due to tahe efficacy of a butyrophenone (haloperido) in the treatment of the Motor-Verbal Tic Disorder. The indications are that Motor Tic Disorders probably respond similarly, but this has not been well demonstrated. Further, the social and functional implications for both sets of symptoms are so different, the Motor-Verbal Tic Disorder portending a more serious outlook for the affected individuals, that the distinction seemed warranted on this basis alone.

Pharmacological Treatment

The only condition of the Stereotyped Movement Disorders for which drug treatment is documented is the Tourette's Tic Disorder, haloperidol being the treatment of choice. The less pervasive and less severe Motor Tic Disorders and the Other Stereotyped Movement Disorders have no established drug treatment. As noted, haloperidol may be effective in them too, though the risk of tardive dyskinesia begets a certain reluctance to use it.

SPEECH DISORDERS NOT ELSEWHERE CLASSIFIED

The awkward modifier not elsewhere classified is necessitated by the fact that some speech and language disorders are listed under Specific Developmental Disorders. There are two disorders included here, Stuttering and Elective Mutism.

At this time, it is not clear whether Elective Mutism is a syndrome in and of itself, or a symptom occurring in a variety of clinical contexts. It was felt that the condition had sufficient distinctiveness to warrant its inclusion as a category. It is possible that research may not bear out this judgement.

Psychopharmacological Treatment

This class of disorders has no known appropriate pharmacological treatment.

CONDUCT DISORDERS

This category is for children who display antisocial behavior and a lack of concern for social norms.

Diagnosis

Two broad classes are usually observed, one consisting of children who have not developed adequate peer relationships for their age and whose antisocial behavior is not usually performed as part of a peer group activity (Undersocialized Conduct Disorders), the other including children with active social involvement antisocial behaviors typically occur in conjunction with a delinquent peer group (Socialized Conduct Disorder). The term under-socialized has been preferred in DSM III over the traditional epithet unsocialized used in DSM II since the latter implies a total absence of socialization which was felt to be overly categorical.

The major difference between the DSM III and DSM II class ifications of the conduct disorders is the differentiation in the former between the aggressive and nonaggressive forms of the undersocialized forms of the disorder. It was felt that a diagnostic distinction should be provided between children with conduct disorders who are aggressive, and those who show no overt aggression. The presence of violent behavior may have distinct implications for the long-term outcome of children with conduct disorders as well as for their pharmacotherapy.

The clinical criteria for the Socialized Conduct Disorder of DSM III stipulates that just deviating from social norms is insufficient. The DSM III diagnosis requires that, in addition to socially disapproved behavior with a peer group, youngsters must also dispay a variety of

dysfunctions (such as relationship difficulties at home and school) to qualify for this coding. Therefore, delinquent behavior alone is not considered a mental disorder in DSM III but can be noted in a section called Conditions Not Attributable to a Known Disorder, Childhood or Adolescent Anti Social Behavior.

Pharmacological Treatment

There is no established pharmacological intervention in the management of conduct disorders, though because of laxity in current diagnostic systems, the evidence is difficult to disentangle from that relating to the Hyperkinetic Reaction. There have been speculations that some adolescent conduct disorders represent later manifestations of Attention Deficit Disorders which, with time, have become complicated with antisocial behavior. In such adolescent cases, the use of stimulants has been reported to be therapeutic. However, this drug effect is far from substantiated as is that of lithium in the Aggressive Conduct Disorder. At this time, the most accurate statement concerning the usefulness of pharmacological treatment in pure conduct disorders uncomplicated by Disorders of Attention or, in former parlance, hyperactivity, is that no such treatment has been demonstrated to be clinically efficacious, though more study based on properly honed diagnosis is highly desirable since indications that some drugs such as stimulants may be useful need confirmation.

EATING DISORDERS

Diagnosis

The single diagnosis of feeding disturbance (in special symptoms for DSM II) has been expanded to include several entities in DSM III: Anorexia Nervosa, Pica, Rumination, Bulimia and other unspecified. The greater diagnostic distinctions found in DSM III are reflective of the pre-Kraepelinian-like splitting approach described earlier.

Pharmacological Treatment

There is no demonstrated effective treatment for any of the eating disorders. An investigation of the effects of Periactin has not shown this drug to contribute significantly to weight gain in women with Anorexia Nervosa (Goldberg et al. 1977), Chlorpromazine is used commonly in Anorexia Nervosa (Dally, 1969), though its combination with bedrest, insulin, contingency management and/or psychotherapy makes elucidation of its therapeutic role difficult.

ANXIETY DISORDERS

Three anxiety disorders specific to childhood are included in DSM III: Overanxious Disorder, Separation Anxiety Disorder, and Shyness Disorder. The distinctions between these are important. For example, the Overanxious Disorder would not include cases of school refusal. The DSM III diagnosis of Overanxious Disorder is for children with excessive, pervasive worry and fearfulness not related to specific events or situations.

The Separation Anxiety Disorder represents a refinement of the overall phobic category. In DSM II, children with abnormal separation reactions could have been diagnosed as neurotic, having a behavior disorder (Overanxious Reaction) or Adjustment Reaction.

Psychopharmacological Treatment

The use of antianxiety agents such as the benzodiazepines has been reported in several clinical studies which claim that their findings support the clinical efficacy of antianxiety medication in children. However, the reports often defy a clear identification of the diagnostic characteristics of the children treated. For example, the samples are often described as neurotic, a term which does little to communicate clinical inclusion criteria. The studies

which have advanced claims of efficacy for antianxiety agents suffer from so many shortcomings that it is not possible to draw any reliable information from them (Gittelman-Klein, 1977).

The efficacy of drug treatment of a clinical subgroup of the childhood anxiety disorders consisting of children with pathological separation anxiety has been studied and the antidepressant imipramine was found to be markedly superior to placebo, in one study from one center (Gittelman-Klein, 1977). This work is in part responsible for delineating the syndrome as separate from other childhood anxiety disorders—an example of how progress in psychopharmacology may influence psychodiagnostic concept.

OTHER DISORDERS OF CHILDHOOD OF ADOLESCENCE

Diagnosis

The three disorders in this rubic do not fall logically into any of the above classes of conditions and do not represent a clinically homogeneous subgrouping.

The introverted disorder refers to children who are loners and who have introverted interests. They typically have no friends and lack social interest in general. These youngsters have been referred to, in the past, as having "shut-in" personalities. In the DSM II, these children have been lumped together with shy, anxious children who are reluctant to initiate social contact, but who enjoy peer interactions once these are established. Distinguishing asocial, isolated children from shy ones will make it possible to determine whether the two diagnoses have different associated treatments and long-term outcome.

Oppositional Disorder includes children who are pervasively negativistic and oppositional in their interaction with authority figures, but who, unlike the Conduct Disorders, do not display marked antisocial behavior. Whether this type of behavior represents a distinct clinical entity, or within the behavior occurs as part of a variety of disorders is unclear.

The last disorder of childhood, Academic Underachievement Disorder, is for children of normal, or above normal academic competence, who because of emotional conflict, fail to perform. They are children traditionally referred to as underachievers. It is questionable whether this category, as well as some others discussed, represents a discrete syndrome or whether the dysfunction is one clinical aspect of a variety of conditions. Its inclusion in DSM III stems from informal reports by clinicians that a pattern of under-achievement in the absence of other psychopathology (particularly specific developmental disorders) is encountered encountered among practitioners who treat middle class children, frequently the offspring of well-educated professional parents.

Pharmacological Treatment

No definite statements can be made, not surprisingly in view of the uncertainty of these entities.

Diagnostic Groups Usually Originating in Childhood But Not So Classified In DSM III

There was considerable discussion about the inclusion of disorders of Gender Identity, or of certain Sleep Disorders under Childhood Disorders. However, in part because of organizational issues, and in part because these disorders do not in fact occupy a significant place in the clinical work of child psychiatrists, these disorders remain under other headings. However, it should be noted that under Psychosexual Disorders, 302.61 specifies Gender Identity or fole disorder of childhood while under sleep Disorders are both Somnambulism and Night Terrors which both commonly arise in childhood and are more common during that age period as well. The latter disorders may be of interest to pediatric psychopharmacologists as the benzodiazepines have proved of value in these disorders.

Specific Developmental Disorders

The Specific Developmental Disorders reflect conditions due to deviations from levels of function expected to occur in children, given usual opportunities for growth and development. These disorders are noted on Axis II; the functions selected are those which are felt to have a potentially disruptive effect on a child's general ability to cope with usual task demands, especially in school. These areas of development include reading and arithmetic skills, language, speech, motor coordination and control of elimination. They are referred to as specific because they can occur in isolation without any other clinical concomitants, though children with a variety of psychiatric disorders are more likely to suffer from Specific Developmental Disorders.

Enuresis and Encopresis are considered simply deviations from normal childhood development and not necessarily part of other, more pervasive clinical disorders. They are therefore included among the Specific Developmental Disorders to be coded on Axis II. A distinction is made between the primary form in which the individual has never developed bladder or bowel control, and the secondary form in which, after a period of continence, loss of elimination control occurs. A rating is made only when there is no known organic abnormality causing the disorders.

A distinction is made in DSM III between speech or articulation difficulties and language or communicative disorders. Most likely articulation and language difficulties are the result of different neurophsyiological disorders; clinically they have very dissimilar consequences and call for different interventions. Therefore, distinguishing between them appears reasonable.

Pharmacological Treatment

Except for the Enuresis, none of the other Specific Developmental Disorders has a relevant chemotherapy. The symptomatic effectiveness of tricyclic antidepressants like imipramine in enuresis is well documented, but there is some question whether it is ever curative and other treatments, such as "bell and pad" conditioning, may be the treatment of first choice.

CAVEAT

There are many substantive differences between the DSM III and DSM II classifications. DSM III followed explicit guidelines which favored splitting rather than lumping disorders together. In addition, there was a policy to include a diagnostic category if it generated clinical interest. There is little doubt that not all such innovations will withstand the test to time. However, if some do, then the field of psychodiagnosis will be rewarded by the approach.

There are some childhood disorders which have specific drug treatment responses, for instance the Attentional Deficit Disorder with Hyperactivity, and possibly the Separation Anxiety Disorder. However, even among those groups there are children who, though they may fit the operational criteria for these disorders, fail to respond to the usual pharmacologic compounds. A lack of response to treatment among these children should not be construed necessarily as challenging the accuracy of the diagnostic assignment. It is likely that each of the childhood disorders identified has multiple etiologies, and the clinical picture may be a final common outcome of diverse pathophysiologies and social antecedents. Consequently, even among well-diagnosed groups of children, one can expect that some individuals will not conform to the established treatment effects and these differences offer the possibility of further valid subclassifications.

It is important that the proposed DSM III classification be viewed as a working tool, one which will need alterations and refinement, and not as a set of fixed entities. Research findings in psychopharmacology already have influenced some aspects of psychiatric diagnosis. It is hoped that future knowledge in psychopharmacology will contribute further to the validity of the pediatric nomenclature.

TABLE 1

DSM II AND III CLASSIFICATIONS FOR
DISORDERS USUALLY ARISING IN CHILDHOOD OR ADOLESCENCE

DSM II (Taken from all categories where mention of childhood)	DSM III (as of March, 1977)
Mental Retardation	**Mental Retardation**
310 Borderline	317.0 Mild
311 Mild	318.0 Moderate
312 Moderate	318.1 Severe
313 Severe	318.2 Profound
314 Profound	319.0 Unspecified
315 Unspecified	
Special Symptoms	**Pervasive Developmental Disorders**
306.00 Speech disturbance	299.00 Infantile autism
306.10 Specific learning disturbance	299.80 Atypical childhood psychosis
306.20 Tic	299.1 Disintegrative Psychosis
306.30 Other psychomotor disorders	299.20 Pervasive developmental disorder of childhood, residual state
306.40 Disorders of sleep	299.90 Unspecified
306.50 Feeding disturbance	**Specific Developmental Disorders**
306.60 Enuresis	Note: These are coded on Axis II
306.70 Encopresis	315.00 Specific reading disorder- Alexia Developmental Dyslexia
306.90 Other special symptoms	315.10 Specific arithmetical disorder
Transient Situational Disturbances	315.30 Developmental language disorder
307.00 Adjustment reaction of infancy	315.40 Developmental articulation Disorder
307.10 Adjustment reaction of childhood	315.50 Coordination disorder
307.20 Adjustment reaction of adolescence	307.6 Enuresis
Behavior Disorders of Childhood and Adolescence	307.7 Encopresis
308.00 Hyperkinetic reaction	315.60 Mixed
308.10 Withdrawing reaction	315.80 Other
308.20 Overanxious reaction	315.90 Unspecified
308.40 Unsocialized aggressive reaction	**Stereotyped Movement Disorders**
308.50 Group delinquent reaction	307.21 Transient Motor Tic disorder
	307.22 Chronic Motor tic Disorder
	307.29 Unspecified tic disorder
	307.30 Other

308.90 Other reaction

Schizophrenia

295.80 Childhood

Speech Disorders Not Elsewhere Classified

307.00 Stuttering
307.91 Elective mutism

Conduct Disorders

312.00 Undersocialized conduct Disorder
307.23 Tourette's disorder
307.29 Unspecified tic disorder
307.30 Other

Speech Disorders Not Elsewhere Classified

307.00 Stuttering
307.91 Elective mutism

Conduct Disorders

312.0 Undersocialized conduct disorder, aggressive type
312.1 Undersocialized conduct disorder, unaggressive type
312.2 Socialized conduct disorder

Eating Disorders

307.10 Anorexia nervosa
307.51 Bulimia
307.52 Pica
307.53 Rumination
307.58 Other or unspecified

Anxiety Disorders

309.21 Separation anxiety disorder
313.20 Shyness disorder
313.00 Overanxious disorder

Disorders of Late Adolescence

309.22 Emancipation disorder of adolescence or early adult life
313.60 Identity disorder
309.23 Specific academic or work inhibition
313.XX Introverted disorder
313.50 Oppositional disorder
313.70 Academic Underachievement disorder

Other Disorders Commonly Diagnosed in Childhood

Sleep disorders

307.46 Somnambulism
307.47 Night terrors

Psychosexual Disorders

302.61 Gender identity or role disorder of childhood

Adjustment Disorders

300.40 with depressed modd
309.28 with anxious mood
309.24 with mixed emotional features
309.82 with physical symptoms
309.30 with disturbance of conduct
309.40 with mixed disturbance of emotions and conduct
309.83 with withdrawal
309.90 other or unspecified

REFERENCES

Dally, P. (Anorexia Nervosa). London, William Heinmann, 1969.

(Diagnostic and Statistical Manual of Mental Disorders, II). Washington, D.C., American Psychiatric Association, 1968.

GAP Psychopathological Disorders in Childhood: Theoretical Considerations and a Proposed Classification. New York, Group for Psychiatry, 2966.

Gittelman-Klein, R. Psychopharmacological treatment of anxiety disorders, mood disorders, and tic disorders of childhood. In M. Lipton (Ed.), A Review of Psychopharmacology: A Second Decade of Progress. New York, Raven Press, 1978, pp. 2471-1480

Goldberg, S., Halmi, K., Eckert, E., Caster, R., & David, J. Cyproheptadine (Periactin) in anorexia nervosa. Paper presented at the annual meeting of the American Psychiatric Association, Toronto, Canada, May 1-7, 1977

Grossman, H.J., Warren S., Begab, M., Eyman, R., Nihaira, K., and O'Connor, G. Manual on Terminology and Classification in Mental Retardation. Baltimore, Maryland, American Association, for Mental Deficiency, 1973.

Kanner, L. Autistic disturbances of affective contact. The Nervous Child, 1943, 1, 217-250.

Klein, D.F. Delineation of two drug-responsive anxiety syndromes. Psychopharmacologia, 1964, 5, 397-408.

Klein, D.F. Drug therapy as a means of syndromal identification and noseological revision. In J. Cole, A. Freedman, A. Friedhoff (Eds.), Psychophathology and Psychopharmacology. Baltimore, The Johns Hokins Press, 1973, pp. 143-160.

Klein, D.F. and Davis, J. Diagnosis and Drug Treatment of Psychiatric Disorders. Baltimore, Maryland, Williams and Wilkins, 1969.

Quay, H. Classifications and patterns of personality. In H. Quay and J. Werry (Eds.), Psychopathological Disorders of Childhood (2nd ED.). New York, Wiley, in press.

Rutter, M., Shaffer, D. and Shepherd, M. An evaluation of the proposal for a multi-axial classification for child psychiatric disorders. Psychological Medicine, 1973, 3, 244-250.

Rutter, M., Shaffer, D. and Sturge, C. A Guide to a Multi-Axial Classification Scheme for Psychiatric Disorders in Childhood and Adolescence. London, Frowde & Co., 1976.

Wender, P. Minimal Brain Dysfunction in Children. New York, Wiley, 1971.

Werry, J. The childhood psychoses. In H. Quay and J. Werry, (Eds.), Psychopathological Disorders of Childhood, 2nd Edition. New York, Wiley, in press.

APPENDIX IV

BEHAVIOR OBSERVATIONS AND ACTIVITY MEASURES FOR USE IN PEDIATRIC PSYCHOPHARMACOLOGY*

I. BEHAVIORAL OBSERVATIONS

Definition

For the purposes of this discussion, behavioral observations will be defined as any procedure in which behavior is measured contemporaneously or as it happens rather than as an algebraic summation of many observations accruing over a long interval of time which is the essence of the more commonly used rating scale or scales of improvement. This technique is distinguished from performance measures in that observations are made on spontaneously occurring behavior rather than behavior evoked under special circumstances such as laboratory or test conditions. A further distinctive feature is that the observation technique conceptualizes behavior within simple categories of externally observable events rather than in derived hypothetical constructs such as intelligence, perception and so on. All these distinctions are, however, relative rather than absolute.

Present status of observational methods

Direct observations of behavior are far from novel. They have long formed part of scientific investigation as in biology, anthropology, psychology and in applications in industry but their use in pediatric psychopharmacology, like the field itself, is new and this method lags in popularity well behind rating scales and performance measures (Conners, 1972). Interest in behavioral methods arises from at least two sources. First, rating scales are vulnerable to observer bias since the observer is usually an interested party such as parent or teacher and the ratings are ordinarily cast in terms of social value judgments such as hyperactivity, aggressiveness and so on (which antipsychiatric critics have been quick to note). But investigators themselves have also been dissatisfied with their reliance on rating scales (Alderton and Hoddinot, 1969, Conners, 1972, Werry and Quay, 1969). However, there is little doubt that the main impetus to the use of behavioral observations in pediatric psychopharmacology has come from the upsurge of a new therapeutic modality for children called behavior modification. Indeed, most of the information an investigator needs to know about behavior observations can be found in a single journal, the Journal of Applied Behavior Analysis, now only in its eighth volume. It may be noted in passing that a further advantage of this journal is that in behavior no doubt well shaped by their own methodology, the editors of this journal maintain a cumulative index with the heading of interest here being "Recording and Measuring Techniques." This should be regarded as one of the standard reference works on the subject and updating the present review can be made relatively quickly by inspecting this cumulative index.

Nevertheless, the method of behavior observation remained an infrequent measurement method of pediatric psychopharmacology and only a few instances are readily citable at this time (Alderton and Hoddinot, 1969; Douglas, 1974; Ellis et al, 1974; Rapoport et al, 1971, 1974; Sprague et al, 1970; Werry and Quay, 1969; Werry and Sprague, 1974). The failure to

*John S. Werry, M.D.

utilize behavior observations in pediatric psychopharmacology lies no doubt at least in part in the logistical cumbersomeness of the technique but the major resistance surely must lie in the less than commendable slowness with which child psychiatry and pediatrics have incorporated the behavioral approach within their technical knowledge.

Potential role in pediatric psychopharmacology

Like any other measurement method in pediatric psychopharmacology, behavior observations have two potential roles. The first is as a diagnostic or other criterion variable (e.g. Douglas, 1974) and the second, rather more common, is as a dependent variable measure of drug effect (e.g. Rapoport et al, 1971, 1974; Sprague et al, 1970). It is obvious that the first, diagnostic or predictor variable, will ordinarily require some kind of normative contrast data while the second does not and hence is simpler to use particularly where the methodologically (for controlling for error) and ethically highly desirable crossover or within subject type of design is used. Such experimental design matters will not be discussed further here since they are common to any measurement technique. Rather it is wished only to emphasize that the behavior observation should be thought of as having a potential role greater than simply that of a measure of drug effect.

As noted above, the particular value of behavior observations should lie in their objectivity and relative freedom from observer bias. But they have also another value characteristic of the behavioral approach in general namely, relevance, in that their meaning is obvious and having been derived from the naturalistic situation, can be seen to be relevant to the kind of complaints made about children by parents, teachers and other caretakers. Noise, running around the classroom, speaking out of turn, and so on are clearly a great deal more meaningful to child caretakers than hypothetical constructs like anxiety or poor self image. In addition, behavioral observations are heuristic in that they take investigators out of their offices and laboratories into the naturalistic situation which has resulted in substantial advances in therapeutic techniques, particularly along problem oritented lines and in understanding of the difficulties faced by parents and teachers in dealing with disturbed children. For all these reasons, then, behavior observations in pediatric psychopharmacology are surely to be encouraged as suggested by Conners (1972) though, like most other measurement techniques, they also have their own set of serious drawbacks.

Disadvantages

The chief problem with this method lies in its logistical clumsiness. Most investigators draw their subjects from a wide area and travel to the home or school is a time consuming business. Additionally, there is often considerable administrative and individual resistance to the introduction of observers into the child's various natural environments, especially the school where the observer may be very threatening to the teacher or principal. There are also likely to be disruptions to observing through class trips, sickness, teacher change, absences and so on. As an alternative, children may be observed in the clinic (e.g. Rapoport et al, 1971) but to do so greatly restricts one of the unique advantages of the behavior observation technique which is its naturalistic nature.

Further disadvantages of this method relate to the problem of sensitivity to drug effect discussed below and its possible reactivity or distortions of the child's behavior and environment usually in the direction of normalising it through the act of observing (Johnson and Bolstad, 1973, 1975).

In summary then, behavior observation method would appear to have sufficient theoretical advantages to make it a desirable part of any standard measurement in pediatric psychopharmacology but because of its peculiar difficulties, its role should be seen as complimentary and validating of other methods rather than a substitute for well established rating scales.

Scope

As with any measurement technique, the behavior observation method must begin by defining what is to be measured. To the extent that this is well done in clear objective and meaningful terms so is the possibility of the measure's usefulness enhanced. Such an initial process requires considerable clinical skill and experience which cuts across any particular theoretical orientation.

Behavior observations may be idiographic or custom designed to suit one particular child's symptoms. This is highly characteristic of much of the behavior modification literature. While these idiographic methods could be well suited to assessing the effect of drug therapy in a particular child once a drug is marketed, it is hard to see how in any initial investigation of a drug, anything other than a nomothetic or general type of method could be useful. Fortunately, the number of behaviors which bring a child to the attention of psychiatrists and pediatricians and for which medication might be indicated are finite and several instruments subsuming most of these behaviors already exist. For example, for classroom work a scale developed by Becker and his colleagues and refined by Werry and Quay (1969) covers most common behavior problems seen in the classroom. It is important that a good scale should also pay attention to positive behaviors such as attention to work, positive interaction and not fall into the trap of just noting bad behaviors. This is to avoid the situation in which a drug, particularly of a depressant type, surpresses bad behavior at the expense of normal function, a good example of which can be seen in Sprague et al (1970) where thioridazine reduced the frequency of deviant behavior but also reduced the level of positive interaction between teacher and child.

While a good general scale should be suited to most types of behavior disorders and to the majority of children for whom drugs are likely to be indicated it is probably unlikely that it should suit different environments. Though for most purposes the general type of scale would be the most useful, there will be situations in pediatric psychopharmacology where more restricted scales may be required. For example, in the treatment of enuresis or Gilles de La Tourette syndrome where simple counts of the symptom would certainly be necessary. Even then it would still be important to use the general scales as well, as a check for oversedation or unexpected behavior change as has been noted for example, to occur in the drug treatment of enuresis (Werry et al, in press (a)).

Academic behavior requires special mention. The behavior modification literature is replete with ways of measuring academic output on an hour by hour basis rather than the traditional achievement tests which are useful only for intervals of several weeks or months. Most of the behavioral methods, however, require programmed or special curriculum materials which would enable quantification of output in unit time. While academic output would be a highly desirable measure of drug effects particularly since extravagant claims have been made for drug effects on learning, the observation method will in most instances require too much interruption of classroom procedures with too much variation between individual classrooms and hence between children to form a useful method in this area. However, the simple noting of time devoted to academic work, attention and so on are easy and part of existing scales including those recommended below. Caution must of course, be exercised about equating such behaviors with actual learning.

Techniques of behavior observing

Methods of making behavior observations are remarkable few.

1. The running of continuous record

   Here all behavior over a given period of time is recorded. Generally speaking, this is unsuitable for anything except automated techniques of recording such as those used for motor activity (Montagu, 1975; Montagu and Swarbrick, 1974, 1975) noise level (Montagu, 1975; Schmidt and Ulrich, 1969) or the familiar tape recording of videotaping

of behavioral sequences. Apart from these instances, the general running record particularly of an ancedotal nature will be restricted to initial open studies when the particular effects of a drug are unknown or, alternatively, in the development of new scales. This is because under these conditions quantification is less important than qualification.

Behaviors which are particularly suitable for automated running records are activity (see special section), noise (Montagu, 1975), urination (Azrin et al, 1971, and the well known Mowrer bedbuzzer) repetitive work (Tate, 1968) and of course, various physiological functions outside the scope of this review. Most automated recording requires complicated equipment and place considerable constraints upon the naturalistic situation and are therefore unsuitable for general use but could from part of an investigator's own extension of a standard battery.

An exception to the above as far as the usefulness of a continuous record is concerned is one where the behavior in question is of low frequency. For example, severe temper tantrums or enuresis. Here a running record would be kept but all that is required is usually a calendar or some such on which the observer concerned notes the occurrence of the event. This technique is simple and accepted but could be more widely used outside traditional symptoms.

2. Time sampling

Where behaviors are frequent it is not possible to record their every occurence, certainly not over any extended time. Under these conditions which are typical of the behavior disorders for which children receive medication, the child's behavior sampled. Typically a period of time thought to be long enough to be characteristic, say half an hour, is devoted to observing the child and the period broken down into sub units of ten to thirty seconds during which the behaviors are noted as occurring or not occurring. In the reviewer's experience, it is highly desirable to have a rest period of say ten seconds for recording and preventing the development of boredom and consequent inaccuracy. There are various mechanical aids to such recording (Schwitzgebel and Ackerland, 1973) such as timers (Foxx and Martin, 1971; Quilitch, 1972), event records, counters and so on detailed in the various behavior therapy journals and advertisements therein but for most purposes a stopwatch or a watch with a good second hand, a pencil and paper are all that will be needed. An example of this method for use in a classroom is described in detail in Werry and Quay (1969).

There are in addition, automated methods of time sampling such as time lapse photography using a simple 8mm movie camera and timing device (Sanders et al, 1969), videotaping, tape recorders and so on again as described in the various behavior therapy journals. But as with continuous recording, most of these require complicated equipment and are not suitable outside special environments such as laboratories or particular classrooms.

How often and how long it is necessary to sample or to achieve what Patterson and Reid (1970) have called the reliability of data sampling, is an empirical question as yet poorly researched. Obviously the sample(s) should be sufficient to give an accurate picture of the whole and the need will vary according to the variability of the patient and his environment(s). Patterson (1969) states that thirty minutes should be enough in the school setting whereas Alevizos and Callahan (1973) found that observations lasting five seconds done twice a day were satisfactory for chronic psychotic patients who had a very low rate of behavioral activity. However, the deciding factors in the end are likely to be more related to logistics and economy than to reliability or validity.

Observers

1. Independent observers

While the ideal may seem to be to use a disinterested observer it carries several disadvantages. First, it increases the cost, second, by introducing a third person it increases the complexity of the observation process, third, it runs the risk of distortion in the child's behavior or social environment (this would seem to be likely where the observer is a stranger who visits for only a short period and where the adult and/or child concerned knows the purpose of the visit. This is not a well researched area and such studies as there are, suggest the effect is less than might be suspected (Johnson and Bolstad, 1973, 1975). All observers require some degree of training and their reliability must be checked at the beginning and sporadically throughout.

2. Participant observers

The literature indicates an increasing use of participant observers, that is, adults who are socially involved with the child and have an interest in has behavioral improvement. Typically, these are parents, teachers, nurses or child care workers. The reasons for this increasing use are largely economic. Evidence is now on hand to indicate that, given proper definition of observational items, reliable and valid data can be obtained by participant observers (e.g. Hall et al, 1971, 1972; Rapoport et al., 1974). However, as with independent observers, reliability cannot be assumed on the basis of somebody else's work and may be quite unsatisfactory especially where the frequency of recording is too disruptive to the observer's routine, where the object of the exercise has not been adequately explained and feedback of its utility is lacking, where some unpleasant contingency for reporting lack of improvement, e.g. school attendeance as found by Schnelle (1974) and where the contact between the investigator and the observer is remote and authoritarian.

In summary then, there is no a priori objection to using participant observers and much in favor logistically and economically in so doing but success will depend on treating the parent or teacher with the same degree of equality and giving them the same amount of training and surveillance as when using paid independent observers, seeing that observing required is compatible with any existing duties of the observer, and providing recording materials which are readily accessible to the observer.

3. Automated observing

This has already been discussed and the opinion offered that most of these are unsuitable for routine use but should be considered when circumstances permit. Careful perusal of the behavior therapy journals (such as Journal of Applied Behavior Analysis, Behavioir Reserach and Therapy, Journal of Behaviior Therapy and Experimental Psychiatry and Behavior Therapy) will provide most of the source data and new developments.

Reliability

This topic has been exhaustively reviewed recently by Johnson and Bolstad (1973) which review is strongly recommended as a standard reference.

Interobserver reliability of most behavior observation techniques is high, averaging well over 80% agreement which is the usual method of computing the liability. This is done by dividing the number of agreement by the total number of paired observations and multiplying by 100. However, recent evidence suggests that this method may grossly overinflate reliability, especially where behaviors are very frequent or very infrequent, and that a more accurate method is to compute two scores, 1. the score interval agreement in which intervals in which neither observer recorded the behavior are discarded and only intervals in which both

observers recorded the presence of the behavior is counted as an agreement (Hawkins and Dotson, 1975); 2. the unscored interval agreement in which agreement is counted only when both observers recorded the absence of the behavior. A disagreement is counted when one observer records the presence of behavior and the other its absence. Intervals in which both observers scored the behavior are ignored.

Doubts cast by Hawkins and Dotson about the accuracy of the % agreement score mean that much of the established reliability of behavior observations needs, to be re-established. Not only has existing literature been subject to this error, but has generally failed to take account of a phenomenon known as instrument decay or observer drift (Johnson and Bolstad, 1973). While interobserver reliability may be good during the training period, there is what looks like an inexorable drift toward inaccuracy though the greatest drift is actually a jump occurring immediately after training period (Johnson and Bolstad, 1973) so there is merit in continuing checks though even this does not seem to deal with the drift between training and beginning work. As might be expected, these checks are likely to be more valid if they are random and covert (Johnson and Bolstad, 1973; Romanczyk et al., 1973). While covert checks are increasingly contrary to the modern industrial ethos, it would seem that given the nature of the problem, observers should expect such as part of their terms of employment. Dealing with parents, teachers and nurses is a more delicate matter though, in the reviewer's experience, careful explanation to parents and teachers can obviate most difficulties. (Nurses prove more obdurate). For example, there is usually little difficulty in getting parents and teachers to accept the double blind placebo control technique which is quite analogous to the covert observer reliability check.

While behavior observations are probably freer to observer bias than some of the other methods such as rating scales, they too, can suffer from this problem in similar ways (Johnson and Bolstad, 1973; O'Leary et al, 1975), another reason for interobserver initial and spot reliability checks.

The most important point about reliability is that while a method of behavior observation may have been established to be reliable by other investigators, no individual observer can be assumed to be reliable and must be demonstrated to be so both before collecting usable data and intermittently throughout his work contract.

Validity

Here the term will be restricted to the question of what the measure actually measures. In the first instance, validity specific to the observing situation will be discussed. It is very obvious that behavior observations carry the least risk of offending validity provided the names of behavioral categories are kept as descriptive as possible with a minimum of inference. For example, out-of-seat behavior in classroom observational methods is much preferable to hyperactivity as a descriptive term. The face validity of such observations when done in naturalistic settings is obvious and the chief concern of the investigator should be ensuring that the observer is reliable and that the realiability of data sampling is adequate. However, validity has a generalised as well as a specific component. Measures are usually assumed to extend in meaning well beyond the particular situation in which the data was recorded. Any relationship to hypothetical constructs like global improvement, anxiety or hyperactivity cannot be assumed for behavioral observation methods and particularly when they are used for diagnostic purposes rather than as dependant variables of drug effect. Fortunately, the purpose of pharmacotherapy in children is mostly symptom rather than cure oriented and hence observations at a behavioral level are highly relevant to measuring and evaluating drug effects. Their validity of course, can be checked against more conventional scales like parent rating scales or vice versa. There has been surprisingly little of this cross validation, though what there has is generally supportive of mutual validity (Abikoff et al, in press; Camp and Zanet 1974; Douglas, 1974; Rapoport et al, 1974). More investigation of cross validation is, however, badly needed.

One last issue in validity has already been alluded to, namely the question of reactivity of the instrument of distortion of behavior by the act of observation. Though in theory there are many possible problems, such little empirical data as there is suggests that it is less of a problem than one might anticipate (Johnson and Bolstad, 1973, 1975).

Sensitivity

This is of course, really a particular kind of validity. However, because the field of pediatric psychopharmacology is a graveyard of promising measures which failed to detect drug effects, sensitivity does require special mention. The basic problem as it seems to this reviewer is that in most instances, drug effects are small and particularly small compared with the normal variations in a child's behavior across time or across environment. Evidence of this can be seen in a study by Werry and Quay (1969) where most of the individual behavioral items carried large variances even when arranged over several days. It is against this background noise or variability that drugs have to work. One advantage of the rating scale is that the score is an algebraic summation of many observations at different times on different days averaged unconsciously by the rater. It is thus no wonder that a measure like the clinical global impression which not only averages over time but averages over behaviors and over situations has consistently proven amongst the most sensitive of measures of drug effect (Lipman et al, 1965) while theoretically much more informative and objective measures have proven insensitive to drug effect.

The interaction between the smallness of drug effects, the magnitude of psychobiological fluctuations and the difficulty of getting adequately large samples of behavior is the single biggest weakness of behavior observations and one which may seriously limit the usefulness of this technique even though as a technique it is not alone in this respect. While observational measures have proven drug sensitive in some instances (Alderton and Hoddinot, 1969; Rapoport et al, 1971, 1974 (parent diaries only); Sprague et al, 1970; Werry and Quay, 1969) others, including the same investigators using previously sensitive techniques have failed to detect drug effect (Ellis et al, 1974; Rapoport et al, 1974 (playroom observations); Werry and Sprague, 1974; Werry et al, in press (b)).*

In summary then, there is probably enough evidence of sensitivity to encourage persistence with this type of measure particularly in view of its other advantages.

RECOMMENDED PROCEDURES

Only scales of a general type will be discussed here. They will be described in terms of the observing situation and subclassified by the type of observer required (participant or independent). In general, the choice of measures has been dictated by (in order of importance):

1. Demonstrated drug sensitivity
2. Demonstrated reliability
3. Universality for likely drug populations of children
4. Universality of observing situations and available observers
5. Simplicity and clarity

*Free field techniques have an added problem of reduced drug effects (see section II).

Inevitably there is the reviewer's own experiences, preferences and knowledge of originators' methods has influenced these choices somewhat particularly where several good instruments are available for one observing situation such as the classroom. Individual techniques are described as follows.

Clinic

1. Playroom (Rapoport et al, 1971)

   This method requires the setting up of a playroom with a few toys and other activity material always in the same way and marking out the floor in grids, observations carried out through a one way mirror or can be done in the room itself. Reliability is good. Independent observers are used and the measure provides activity and distractibility scores. Data on sensitivity is conflicting with one positive (Rapoport et al, 1971) and two negative (Rapoport et al, 1974; Werry et al, in press (b)). Also validity outside the particular observing situation is uncertain.

2. Psychologist's frequency counts of distractibility and behavior problems during test (Rapoport and Benoit 1975 (see Rapoport et al, 1974)).

   Here the psychologist counts the number of her intrusive responses into the testing situation. No details on the reliability are presently available but this method correlates significantly though not substantially (r = approximately 0.40) with teacher measures of hyperactivity and conduct problems.

Ward or day patient center settings

1. Alderton and Hoddinott (1968)

   This scale rates ten items of aggressive, affectionate behavior and motor activity for a sampling of three minute periods seven times daily. Items are scored simply as present or absent during the three minute period and is reliable and drug sensitive on the only occasion it has been.used (Alderton and Hoddinott, 1968). The authors used independent observers but it would seem that this method would be easily adapatable to be done by nursing staff using spot checks rather than 3-minute observation periods. When independent observers are used the method could probably be improved by frequency counting every ten seconds rather than simply once during the 3-minute period.

2. Monkman (1972) - independent observations

   This is a method of independent observations and is a complex time sampling scale covering staff reaction, verbal, fine motor, gross motor behavior in a two dimensional system in which the direction of the behavior is also noted (towards object, self, peers or staff or group). There is also an 'inert' item. It is highly reliable, sampling is for two minutes at randomly distributed intervals throughout the day and while its drug sensitivity is not known it is responsive to other therapeutic endeavors. It is included because of its suitability for inpatient units, its long use and refinement over a 3-year period, its comprehensiveness and its extensive and clear documentation.

3. Monkman (1972) - daily check list

   This is a list of 26 items covering daily routines common to most inpatient units and checked by nursing or caretaking staff. Some of the items are ratings rather that true behavior observations but most are behavioral events (for example, makes bed, brushes teeth). Realiability is good but sensitivity is unknown.

Play situation (e.g. Summer camps, playground, nursery school, inpatient units)

1. Parten scale (Wintre and Webster, 1974)

   This is an old scale dating from the 1930's revised by two behaviorists who added a seventh item (adult directed activities) to six describing various kinds of play including the wellknown parallel type. Time sampling (at ten second intervals) by independent observers was the method used by Wintre and Webster, though it should be adaptable to spot checking by teachers and other caretakers. Realiability is satisfactory and it is sensitive to psychological intervention techniques though has not been used for drug studies.

Classroom (Werry and Quay, 1969)

1. Werry and Quay (1969) adapted a scale by Becker, O'Leary and others which consists of about ten items relating to deviant behavior, attention and teacher pupil interactions. Behavior is time sampled for twenty seconds followed by ten second rest and the duration of sampling can be adapted to fit the investigator's need. The reliability is high but variances are too, which is likely to make sensitivity a problem. It has shown drug effects twice (Sprague et al, 1970; Werry and Quay, 1969) and failed to do so once (Werry and Sprague, 1974). If necessary it can be simplified by recording just some of the items particularly attention and teacher/pupil interaction. A modification (Zimet, Camp and White, 1977) has shown good correlations with teacher ratings of prosocial behavior but a variable relationship to teacher ratings of negative behavior (Camp & Zimet, 1974).

2. Kubany and Sloggett (1973). This is a method of time sampling by the teacher who determines the interval to suit herself. Kubany and Sloggett used four, eight or sixteen minute intervals for sampling but there would appear to be no reason that it could not be done on a purely random basis at the teacher's convenience. There are three very simple scoring categories - on task, passive behavior and disruptive behavior. The validity appears satisfactory and while it is sensitive to psychological therapeutic intervention it has not been used in drug studies. This would seem to be a simple measure but faces the difficulty of teacher cooperation.

3. Abikoff et al (in press) have adapted a classroom observation scale specifically for hyperactive children consisting of 14 categories recorded at 15 second intervals during seat work over a 32 minute period. While this scale has one defect compared with Werry and Quay's, the absence of prosocial items such as positive child teacher interaction, it has the strengths (in addition to reliability) of demonstrated discrimination between hyperactive and normal children and good cross validation with Conners TQ. However, its sensitivity to pharmacotherapy has only been established in a preliminary way and it seems, a priori, to be so qualitatively similar to that of Werry and Quay that problems with irregularity of sensitivity would also be expected.

Home

1. Parent diary of events (Rapoport et al, 1974).

   Four day diary recording what a child is doing at hourly intervals by one or other parent, usually the mother, and scored by the rater post hoc for two categories, activity level and behavior problem. Reliability is not stated but there are low but significant correlations with other measures such as psychologist's ratings and it is reported to have been validated against direct observations of behavior in the home. This has the advantages of simplicity and economy of staff time and has been shown to be drug sensitive (Rapoport et al, 1974).

2. Independent observer time sampling (Rapoport and Benoit, 1975).

This procedure counted the frequency of various behaviors (activity change) during a half hour free play session. Interrater reliability was satisfactory and, in that study, the measure was drug sensitive. In general, however, free play measures have not been sensitive to drug effects.

3. Time sampling by independent observer (Hawkins et al, 1966).

Behavior is sampled at ten second intervals along a number of categories but also includes response by parents and others to these behaviors. It is reliable and sensitive to psychological therapeutic intervention but has not been tried with drugs.

4. Family interactions (Patterson et al, 1971).

This is a time sampling procedure for studying family interactions. The observer goes to the home around dinner time and makes two five minute observations of each family measure. However, for pediatric psychopharmacology it could be restricted to observations of the child on medication and his interaction with other family members over a meal time. There are 29 categories of behavior which makes it rather complex but it is the only good family interactional measure available. It has not been used in drug studies and details can be obtained in Patterson et al (1971). It is recommended in the hope that someone may wish to look at this complex but important area of the effect of drug on child/family interactions.

CONCLUSIONS

Behavior observations offer a degree of realiability, objectivity and face validity which is high compared with other diagnostic and dependent variable measures of drug effect. Technology of observing is well established and documented in behavior therapy journals especially the Journal of Applied Behavioral Analysis. Scales are available for various needs and others can be developed as needed to suit particular target populations. The observations may be done in a laboratory or a naturalistic setting and may employ independent or participant observers or, in certain instances, be automated. The weaknesses of the method lie in its cost, ligistical complexity, high variances, uncertain sensitivity to drug effects, instrument decay or reliability drift over time, reactivity to the act of observing and social resistance to allowing observers into home, ward or classroom. Use of the method in pediatric psychopharmacology has so far, been limited, and with mixed success. The method should be regarded as highly desirable, worthy of further application but should probably not be obligatory in any battery of measures except perhaps for some kind of playroom measure at the clinic itself which would require a minimum amount of staff and travelling time. Unfortunately, such a playroom measure would fail to realize the principal advantage of behavior observations which is their relationship to the child's real world.

II. ACTIVITY MEASURES

Definition

Activity will be defined as the sum total of movement of the whole plus any part of the body in space occuring in unit time. As such, it is a quantitative measure of kinetic energy output through the motor system.

Activity as a psychopathological symptom

Complaints about a child's activity are common in child psychiatric populations (Werry and Sprague, 1970). These complaints may be too much activity (hyperactivity) or, less commonly, too little (hypoactivity). Such quantitative judgments assume some hypothetical norm against which any child's activity can be assessed. Since, in the ordinary course of events, these norms are not explicit, judgments of abnormality of activity necessarily must be social value judgments in the majority of instances. The notion that there is an "hyperactive child" in a quantitative sense is deeply entrenched in child psychiatry and behavioral pediatrics and indeed, until recently hyperactivity was seen as the core symptom of the minimal brain dysfunction syndrome (Clements, 1966; Wender, 1971) and the prime indication for the use of psychotropic drugs in children (Academy of Pediatrics 1975). A decade's intensive study of the hyperactive child has, however, produced a shift towards defining the fundamental deficit as of attention rather than of activity (Douglas, 1974) soon to be reflected in both the IXth Revision of the International Classification of Diseases and the Third Diagnostic and Statistical Manual of the American Psychiatric Association.

Conceptual problems

As already noted, implicit in all the work on the hyperactive child and in most of pediatric psychopharmacology, is the notion that there exists as a stable dimension of behavior a mean daily activity level characteristic of any particular child. While this level may be subject to large fluctuations throughout the day, across days and across environments, nevertheless averaged out each child has his own characteristic level. A further assumption is that there is some etiological connection between high activity level on the one hand and brain damage and younger age on the other.

To what extent is this assumption of a mean daily activity level true? Studies carried out up to 1973 reviewed comprehensively by Sprague and Werry (Sprague and Werry, 1971, 1974; Werry and Sprague, 1970) did not lend much credence to the notion and work since then has created even further doubt (Gittelman-Klein and Klein, 1975; Routh et al, 1974; Shaffer, 1973; Shaffer et al, 1974). Most studies have shown little correlation between activity level in one environment and that in another. Neither is there any consistent correlation between high activity level and brain damage (Shaffer, 1973; Shaffer et al, 1974; Werry, 1972; Werry and Sprague, 1970) though the inverse correlation of activity with age has been repeatedly confirmed (Routh et al, 1974; Sprague and Werry, 1971, 1974).

How then can the notion of the hyperactive child persist - which it does even more strongly than before? It was suggested by the reviewer (Werry, 1968; Werry and Sprague, 1970) and subsequently confirmed in empirical studies by Douglas (1974) and Shaffer (Shaffer et al, 1974) that it is the situational social inappropriateness or disruptiveness of the movements which is distinctive. Some children, reasonably persistently, exhibit more movement in specific situations such as the classroom or at home when they are expected to be still. Hyperactivity, then, is a particular kind of conduct disorder which is characterized by (usually) non aggressive movement disruptive to one of the small social systems of which the child is a member.

Should the notion of hyperactivity be abandoned then? Even if the answer to this were to be in the affirmative, the notion is too deeply entrenched to be got rid of easily though, as already noted, there is a shift towards focusing on attention as the primary dysfunction (Douglas, 1974). It would seem more realistic to accept hyperactivity as a distinctive symptom of psychopathology but to recognize what the term implies: a special kind of conduct disorder which is socially disabling to the child and deserving of treatment. A second reason for retaining the concept is that hyperactivity is usually incompatible with learning and is associated often, though not inexorably, with persistent academic retardation (Douglas, 1974; Sprague and Werry, 1971, 1974; Wender, 1971) and it is appropriate to try to ameliorate it (as with psychotropic drugs in the hope that learning may be facilitated). However in so doing one should not be so naive as to assume that learning will necessarily follow automatically (Douglas, 1975; Gittelman-Klein et al, 1978). A rather obvious example is that a sleepy hyperactive child is certainly quieter but is unlikely to be learning more. There is also good reason to suspect that in a significant proportion of cases hyperactivity follows and is dependent on an attentional or other cognitive deficit rather than vice versa and treating the hyperactivity alone is unlikely to do much for learning (Douglas, 1974; Gittelman-Klein and Klein, 1975 Weiss, 1975). Third, there seems little doubt that true or false, the concept of hyperactivity has been one of the more heuristic in child psychiatry in the last decade leading to much interesting theory (e.g. Wender, 1971) and a significant amount of good empirical research.

Methods of measurement of activity

It is not proposed here to restate the methodological issues already discussed at length in Section I on Behavior Observations. Methods may be grouped into rating scales, observations and mechanical or automated techniques.

1. Rating Scales

There are several of these some of which have been discussed elsewhere in this manual and in various published articles (Conners, 1972, 1973). Suffice it to say that most symptom rating scales for children appear to contain an underlying dimension or factor of hyperactivity as can be seen for example, in the widely accepted Conners Teachers Questionnaire (TQ) and Parent Questionnaire (PQ) (Conners, 1972; 1973). This dimension of hyperactivity is usually drug sensitive (e.g. Conners, 1972; Rapoport et al, 1971, 1974; Werry and Sprague, 1974, Winsberg et al, 1972). Interestingly, however, despite the wide acceptance of these two Conners Scales their interobserver reliability and external validity has seldom been tested and then only incompletely (e.g. Gittelman-Klein and Klein, 1975; Rapoport et al, 1974). While it is true that drug produced changes in the hyperactivity factor tend to covary in group averaged data with other measures of activity such as those on the PQ, seat movement, parent and psychologist observations of activity level (Conners, 1972; Rapoport et al, 1974; Sprague and Werry, 1974; Werry and Aman, 1975) there have been few attempts to correlate changes on two independent measures of activity as individual children's scores. This clearly requires further study, particularly since the study by Gittelman-Klein and Klein (1975) shows poor intercorrelation.

A commonly used parent rating scale is the Werry-Weiss-Peters Activity Scale (Werry and Sprague, 1970, Werry, 1968) which, unlike the Conners PQ, is, at least theoretically, a unidimensional scale confined to activity alone. It consists of 22 items each relating to the child's activity in a specific situation such as at meals or watching television. This measure has been used not infrequently in drug studies and has as far as the reviewer is aware, always proven drug sensitive (Conners, 1972; Gittelman-Klein and Klein, 1975; Rapoport, 1971) but, as with the TQ and the PQ though changes tend to covary with those in other activity measures, there have been few efforts directly to intercorrelate it with other activity measures or to look at the reliability of the scale. Routh et al (1974) found that the scale yielded not one activity factor but seven discrete factors, some behavior specific (e.g., verbal behavior) and others situation specific (television

behavior). (This incidentally is a further confirmation of the elusiveness of the activity level concept). Neither was there good correlation between this scale and objective methods of estimateing activity in the laboratory. The scale did, however, show the expected age effects. In one of the few other validating studies, Shaffer et al., (1974) were unable to find any correlation between objective activity measures, such as grid crossing and actometer, and the scale except for a low one between seat activity during the continuous performance test. They suggest that like the TQ and the PQ this scale is more a measure of a special kind of conduct disorder than a true activity measure Gittleman-Klein and Klein (1975) confirmed this. PQ and Activity Scale did correlate with each other but not with non parent derived measures. As with the TQ and the PQ, this scale needs more research into its reliability and validity as an activity measure. In the meantime, it should probably be retained as a measure of parent perceived "activity" largely because of its proven drug sensitivity.

While there are other symptom rating scales containing items or dimensions of hyperactivity, none has the degree of common acceptance in pediatric psychopharmacology of the three mentioned above nor appears to have sufficiently distinctive features to warrant replacing these three.

2. Direct observations

Most of what has been said in Section I on Behavior Observations is particularly germane here. Time sampling is the common method though mechanical methods of continuous recording (see below) are particularly common as measures of activity level. Most of the general behavior scales recommended in Section I already incorporate activity measures or if not, could have an item or two added quite easily. As noted, these scales are appropriate mostly for one particular environment such as the classroom. Where activity is the principal object of the measure, the situation of observing is typically "free field" in a playroom or laboratory and activity is measured by means of grids or other devices marked on the floor dividing the room into quadrants or finer divisions with a selection of toys, etc. scattered throughout the room to promote locomotion. This technique generally requires little equipment and provides an easy way of counting locomotion. (Kalverboer, 1975; Pope, 1970; Rapoport et al, 1971; Routh et al, 1974; Shaffer et al, 1974; Sprague and Werry, 1971; Werry and Sprague, 1970). This is a simple and reliable method and its face validity is obvious but is dependent on the availability of independent observers. However, its drug sensitivity is not always high (see Section I) possibly because drugs may be effective only in situations in which activity is constrained (Ellis et al, 1974) and/or attention required (Douglas, 1975; Werry and Aman, 1975) in experimenter-paced tasks (Sykes et al, 1971). Most of the activity measures also include measures of attention and this is highly desirable in view of the fundamental deficit in this area exhibited by most hyperactive children (Douglas, 1974).

3. Mechanical or automated measures of activity

Several good though often rather expensive and complex techniques are available for both free field and constrained situation (Sprague and Werry, 1971; Werry and Sprague, 1970). One of the oldest is Schulman and Reisman's actometer (see Shaffer et al, 1974), a modified automatic winding wristwatch. This has been used by several investigators and appears reasonably valid and reliable (see Shaffer et al, 1974; Sprague and Werry, 1971; Werry and Sprague, 1970). Its chief weakness lies in its sensitivity to acceleration rather than quantity of movement, the necessity for frequent recalibrating and its uniplanar nature so that movement at right angles to this plane may produce no recording at all. (This can be obviated by using two at right angles to each other). Its usefulness too, is restricted to measuring movement of the part of the body to which it is attached. It is thus most suitable for recording locomotion or hand movement but cannot therefore be really regarded as a true measure of general activity though few other measures except perhaps the ultrasonic method fulfill this criterion. A device rather similar to the actometer is the pedometer (Rapoport et al, 1971; Sprague and Werry, 1971).

The actometer and pedometer have proven to be drug sensitive (Millichap and Johnson, 1974; Rapoport et al, 1971; Sprague and Werry, 1971).

Other free field measures are the ultrasonic method first used by Peacock and Williams for animals (1962) and recently adapted by Montagu and Swarbrick (1974) for children in a playroom situation. This method relies on creating ultrasonic standing waves and measuring their disruption by movement using the doppler effect. Montagu claims to have solved two basic problems which caused the author to abandon this method in 1963; namely the influence of velocity and position with respect to the transmitter on the size of the analog effect. Montagu uses several transmitters and receivers and checks the whole through an object mounted on a gramophone turntable which can be put in various positions and presumably run at various speeds. The equipment and the ways of processing the data accrued are complicated and expensive but it would appear to be a promising laboratory technique for measuring total activity in a free field situation. Further, it appears to be drug sensitive (Montagu, 1975; Montagu and Swarbrick, 1975).

Montagu also uses a grid of pressure sensors in the floor of his laboratory to detect locomotion (Montagu and Swarbrick, 1974) and this measure too, is drug sensitive but would seem to have little to offer (beyond automated recording) over the more traditional grid crossing method described above. Other techniques used in the past for measuring locomotion particularly, have employed grids of light and photoelectric cells (see Sprague and Werry, 1971; Werry and Sprague, 1970).

Also used have been cameras, videotape techniques and tape recorders (Ellis et al, 1974; Kalverboer, 1975; Sprague and Werry, 1971). These require often expensive equipment and place constraints upon use in the child's own environment which is likely to restrict their use in a laboratory and hence sharply limit their distinctiveness and usefulness but may still suit individual investigators.

Restricted field methods include Sprague's stabilimetric cushion, a spring mounted seat with four pressure sensitive switches in each quadrant (Sprague and Toppe, 1966). This is useful for continuous measures of motor activity while the child is seated as for example, in the classroomor while performing a task such as the Continuous Performance Test in the laboratory. In addition to the chair, this method requires only a simple set of counters. In pediatric psychopharmacology, its applicability is probably restricted to the laboratory situation because of the wide dispersion of subjects in most clinical studies. Care must be taken to prevent audible noises coming from the switches or the chair or the counters since these tend to be reinforcing to movement with children. This method has been shown on several occasions by Sprague and the author to be reliable and drug sensitive (Sprague and Werry, 1971, 1974; Sprague et al, 1970, 1974; Werry and Aman, 1975) and would seem to be a useful measure of motor overflow during attentional and other performance tasks. Typically, drugs which improve attention reduce motor overflow too.

Recommended measures

1. Rating scales

   Conners Teacher Questionnaire and Parent Questionnaire, Werry-Weiss-Peters Activity Scale

2. Behavior observations

   The scales by Rapoport et al (1971) or grid crossings (Routh et al, 1974) are suitable without change. Some of the classroom, ward and home measures already include activity measures and where lacking, direct measures of activity could easily have these added as required.

3. Mechanical methods

These should be considered optional since they require specialized equipment which may not be available to all investigators. Of the current methods, Schulman and Reisman's actometer is clearly the simplest and best (see Shaffer et al, 1974 or Millichap and Johnson, 1974 for details). The automated grid and ultrasonic devices of Montagu (Montagu and Swarbrick, 1974, 1975) are all well worth considering for investigators with the necessary resources. Kalverboer's automated laboratory for observing preschool children also deserves some consideration because of its highly developed state for this neglected age group.

Sprague's stabilimetric chair (see Sprague and Toppe, 1966 or write R. L. Sprague, Institute of Child Behavior, University of Illinois, Champaign, Ill. 61820) is highly recommended where laboratory performance tasks are part of the battery since this measure can be easily incorporated.

Summary

Activity level probably does not exist as a stable, quantitative behavioral dimension but consistent, situationally specific inappropriate motor behavior probably does. The core symptom of the socalled hyperactive child probably lies in his inability to control his attention in situations which require him to direct his attention and hence also his motor activity in one particular adult-determined direction. It is possible, however, that activity measures could be of greater importance in the study of preschool or retarded children where motor activity forms a greater part of the total behavior output (e.g. Montagu, 1975; Montagu and Swarbrick, 1975).

The environments chosen should be preferably naturalistic, and socially relevant since there is little generalization of activity across different environments. While laboratory free-field measures are likely to be the most convenient, since they are devoid of the adult directed attentional constraints on movement, they are the least likely to be affected by drugs. The simplest and most reliably drug sensitive methods of measuring activity are still parent and teacher rating scales though their validity as measurements of activity requires to be established. Reliable behavior observation methods are available and should be considered even though they are logistically more complex and subject to the problems (including sensitivity) described in Section I. Activity measures should always include or be complemented by measures of other functions. Good mechanical measures exist for measuring various aspects of motor activity in a free field situation or while seated.

REFERENCES - BEHAVIOR OBSERVATIONS

Alderton, H. and Hoddinott, B.: A controlled study of the use of thioridizine in the treatment of hyperactive and aggressive children in a children's psychiatric hospital. Canadian Psychiatric Association Journal, 1969, 9, 239-24.

Azrin, N., Bugle, C. and O'Brien, F.: Behavioral engineering: two apparatuses for toilet training retarded children. Journal of Applied Behavior Analysis, 1971, 4, 249-253.

Alevizos, P. and Callihan, E.: Observation instruments and the reliability of data sampling in direct patient observation. Paper presented to a symposium on resolving methodological issues of direct behavioral observation in applied settings, presented to the 81st Annual Convention of the American Psychological Association, Montreal, August 1973.

Camp B., Zinet S.G.: The relationship of teacher rating scales to behavior observations and reading achievement of first grade-children, J. of Special Education, 1974, 8, 353-359.

Conners, C.: Pharmacotherapy of Psychopathology in Children in Quay, H. and Werry, J. eds, Psychopathological Disorders of Childhood. New York, Wiley, 1972, 316-348.

Conners, C.: Rating scales for use in drug studies with children. Psychopharmacology Bulletin. Special Issue on Pharmacotherapy with Children. 1973, 24-84.

Douglas, V.: Differences between normal and hyperkinetic children in Conners, C. (ed) Clinical Use of Stimulant Drugs in Children, Amsterdam, Excerpta Medica 1974, 12-22.

Ellis, M., Witt, P., Reynolds, R. and Sprague, R.: Methylphenidate and the activity of hyperactive children in the informal setting. Child Development, 1974, 45, 217-220.

Foxx, R. and Martin, T.: A useful portable timer. Journal of Applied Behavior Analysis, 1971, 4, 60.

Hall, R., Fox, R., Willard, D., Goldsmith, L., Emerson, M., Owen, M., Davis, F. and Porcia, E.: The teacher as observer and experimenter in the modification of disputing and talking out behaviors. Journal of Applied Behavior Analysis, 1971, 4, 141-149.

Hall, R., Axelrod, S., Tyler, L., Grief, E., Jones, F., and Robertson, R.: Modification of behavior problems in the home with a parent as observer and experimenter. Journal of Applied Behavior Analysis, 1972, 5, 53-64.

Hawkins, R. and Dotson, V.: Reliability scores that delude; and Alice in Wonderland trip through the misleading characteristics of interobserver agreement scores in interval recording. In Ramp, E. and Semb, G. eds. Behavior Analysis: Areas of Research and Application Englewood Cliffs, N.Y. Prentice-Hall 1975, 359-376.

Hawkins, R., Peterson, R., Schweib, D. and Bijou, S.: Behavior therapy in the home: amelioration of problem parent-child relation with the parent in a therapeutic role. Journal of Experimental Child Psychology, 1966, 4, 99-107.

Johnson, S. and Bolstad, O.: Methodological issues in naturalistic observation: some problems and solution for field research in Hamerlynck, L., Handy, L. and Mash, E. Behavior Change: Methodology Concepts in Practice, 4th Banff International Conference on Behavior Modification, Champaign, Illinois, Research Press, 1973, 7-67.

Johnson, S. and Bolstad, O.: Reactivity to home observation; a comparison of audiorecorded behavior with observers present or absent. Journal Applied Behavior Analysis, 1975, 8, 181 185.

Kubany, E. and Sloggett, B.: Coding procedure for teachers. Journal of Applied Behavior Analysis, 1973, 6, 339-344.

Lipman, R., Cole, J., Park, L. and Rickels, K.: Sensitivity of symptom and non symptom-focused criteria of outpatient drug efficacy. American Journal of Psychiatry, 1965, 122, 26-27.

Monkman, M.: A milieu therapy program for behaviorally disturbed children. Springfield, Illinois, Charles C. Thomas, 1972.

Montagu, J.: The hyperkinetic child: A behavioural electrodermal and EEG investigation. Developmental Medicine and Child Neurology, 1975, 17, 299-305.

Montagu, J. and Swarbrick, L.: Hyperkinesis: the objective evaluation of therapeutic procedures. Biological Psychology, 1974, 2, 151

Montagu, J. and Swarbrick, L.: Effect of amphetamines in hyperkinetic children:. stimulant or sedative? A pilot study. Developmental Medicine in Child Neurology, 1975, 17, 293-298.

McQueen, M.: A milieu therapy program for behaviorally disturbed children. Springfield, Illinois, Charles C. Thomas, 1972.

O'Leary, K., Kent, R. and Kanowitz, J.: Shaping data collection congruent with experimental hypothesis. Journal of Applied Behavior Analysis, 1975, 8, 43-51.

Patterson, G.: A community mental health program for children. In Hamerlynck, L., Davidson, P., Acker, L. eds. Behavior Modification and Ideal Mental Health Services. 1st Banff International Conference on Behavior Modification, Calgary, Canada, University of Calvary Press, 1969, 130-179.

Patterson, G. and Cobb, J.: A dyadic analysis of "aggressive" behavior in Hill, J. ed. Minnesota Symposia on Child Psychology, 5, Minneapolis, University of Minnesota Press, 1971

Patterson, G. and Reid, J.: Reciprocity and coersion: two facets of social systems. In Neuringer, C. and Michael, J. eds. Behavior Modification in Clinical Psychology, New York, Appleton-Century-Crofts, 1970, 133-177.

Quilitch, H.: A portable programmed, audible timer. Journal of Applied Behavior Analysis, 1972, 5, 18.

Rapoport, J., Abramson, A., Alexander, D. and Lott, I.: Playroom observations of hyperactive children on medication. Journal of the American Academy of Child Psychiatry, 1971, 10, 524-534.

Rapoport, J., Quinn, P., Bradbard, G., Riddle, K. and Brooks, E.: Imipramine and methylphenidate treatments of hyperactive boys: a double blind comparison. Archives of General Psychiatry, 1974, 30, 789-793.

Romanczyk, R., Kents, R., Diament, C., and O'Leary, K.: Measuring reliability of observational data: a reactive process. Journal of Applied Behavior Analysis, 1973, 6, 175-184.

Sanders, R., Hopkins, V. and Walker, M.: An inexpensive method for making data records of complex behaviors. Journal of Applied Behavior Analysis, 1969, 2, 221-222.

Schmidt, G., and Ulrich, R.: Effects of group contingent events upon classroom noise. Journal of Applied Behavior Analysis, 1969, 2, 171-179.

Schnelle, J.: Abrief report on the invalidity of parent evaluation of behavior change. Journal of Applied Behavior Analysis, 1974, 7, 341-343.

Schwitzgebel, R. and Ackerland, V.: Psychotechnology supplement: a compilation of techniques and apparatus for behavior modification. Montclair, California, 91763, Psychotechnology Laboratory (8857 Central Avenue).

Sprague, R., Barnes, K. and Werry, J.: Methylphenidate and thioridazine: learning, reaction time, activity in classroom behavior in emotionally disturbed children. American Journal of Orthopsychiatry, 1970, 40, 615-628.

Tate, B.: An automated system for reinforcing and recording retardate work behavior. Journal of Applied Behavior Analysis, 1968, 1, 347-348.

Werry, J., Aman, M. and Lampen, E.: Methyphenidate and haloperidol in hyperactive children. Acta Paedopsychiatrica, in press(b).

Werry, J., Dowrick, P., Lampen, E., Vamos, M.: Imipramine in Enuresis: psychological and physiological effects. Journal of Child Psychology and Psychiatry, in press(a).

Werry J. and Quay, H.: Observing the behavior of elementary school children. Exceptional Children, 1969, 35, 461-470.

Werry, J. and Sprague, R.: Methylphenidate in hyperactive children; effect of dosage. Australia and New Zealand Journal of Psychiatry, 1974, 8, 9-19.

Wintre, M. and Webster, C.: A brief report on using a traditional social behavior scale for disturbed children. Journal of Applied Behavior Analysis, 1974, 7, 345-348.

REFERENCES - ACTIVITY MEASURES

Abikoff, H. Gittelman-Klein, R. and Klein, D.F.: Validation of a Classroom Observation Code for Hyperactive Children. Journal Consulting Psychology, in press.

American Academy of Pediatrics, Council on Child Health. Medication for hyperkinetic children. Pediatrics, 1975, 55, 560-561.

Clements, S. Minimal brain dysfunction in children. Washington, USPHS NINDE Monograph No. 3, 1966.

Douglas, V.: Differences between normal and hyperkinetic children. In Conners, C.(ed) Clinical Use of Stimulant Drugs in Children, Amsterdam, Excerpta Medica, 1974, 12-22.

Ellis, M., Witt, P., Reynolds, R. and Sprague, R.: Methylphenidate and the activity of hyperactive children in the informal setting. Child Development, 1974, 45, 217-220.

Gittelman-Klein, D.: Are behavioral and psychometric changes related in methylphenidate treated hyperactive children? International Journal of Mental Health, 1975, 4, 182-198.

Gittelman-Klein, R. and Klein, D.: Methylphenidate effects in learning disabilities; I. psychometric changes. Archives of General Psychiatry, 33: 655-664, 1976.

Kalverboer, A.: A neurobehavioral study in preschool children. Clinics in Developmental Medicine No. 54, London, Heinmann, 1975.

Millichap, G. and Johnson, F.: Methylphenidate in hyperkinetic behavior: relation of response to degree of activity in brain damage, in Conners, C. (ed) Clinical Use of Stimulant Drugs in Children. Amsterdam, Excerpta Medica, 1975, 130-139.

Montagu, J.: The hyperkinetic child: A behavioural electrodermal and EEG investigation. Developmental Medicine and Child Neurology, 1975, 17, 299-305.

Montagu, J. and Swarbrick, L.: Hyperkinesis: the objective evaluation of therapeutic procedures. Biological Psychology, 2, 151

Montagu, J. and Swarbrick, L.: Effect of amphetamines in hyperkinetic children: stimulant or sedative? A pilot study. Developmental Medicine and Child Neurology, 1975, 17, 293-298.

Peacock, L. and Williams, M.: An ultrasonic device for recording activity. American Journal of Psychology, 1962, 75, 648-652.

Pope, L.: Motor activity in brain injured children. American Journal of Orthopsychiatry, 1970, 40, 783-794.

Rapoport, J., Abramson, A., Alexander, D. and Lott, I.: Playroom observations of hyperactive children on medication. Journal of the American Academy of Child Psychiatry, 1971, 10, 524-534.

Rapoport, J., Quinn, P., Brandbard, G., Riddle, K. and Brooks, E.: Imipramine and methylphenidate treatment of hyperactive boys: a double blind comparison. Archives of General Psychiatry, 1974, 30, 789-793.

Rapoport, J., Benoit, M. The relation of direct home observations to the clinic evaluation of hyperactive school age boys. J. Child Psychol Psychiatry, 1975, 16, 141-147.

Routh, D., Schroeder, C. and O'Tuama, L.: Development of activity level in children. Developmental Psychology, 1974, 10, 163-168.

Shaffer, D.: Psychiatric aspects of brain injury in childhood: a review. Developmental Medicine and Child Neurology, 1973, 15, 211-200.

Shaffer, D., McNamara, N., and Pincus, J.: Controlled observations on patterns of activity, attention, and impulsivity in brain damaged and psychiatrically disturbed boys. Journal of Psychological Medicine, 1974, 4, 4-18.

Sprague, R., Barnes, K. and Werry, J.: Methylphenidate and Thioridazine: learning, reaction time, activity and classroom behavior in disturbed children. American Journal of Orthopsychiatry, 1970, 40, 615-628.

Sprague, R., Christensen, D. and Werry, J.: Experimental psychology in stimulant drugs, In Conners, C. (ed) Clinical Use of Stimulant Drugs in Children. Amsterdam, Excerpta, Medica, 1974, 141-163.

Sprague; R. and Toppe, L.: Relationship between activity level and delay of reinforcement in the retarded. Journal of Experimental Child Psychology, 1966, 3, 390-397.

Sprague, R. and Werry, J.: Methodology of psychopharmacological studies with the retarded, in Ellis, N. Ed. International Review of Research in Mental Retardation, New York Academic Press, 1971, 5, 147-219.

Sprague, R. and Werry, J.: Psychotropic drugs in handicapped children in Mann, L. and Sabatino, D. (eds) Second Review of Special Education, Philadelphia, JSE Press, 1974, 1-50.

Sykes, D., Douglas, V., Weiss, G. and Minde, K.: Attention in hyperactive children and the effect of methylphenidate (Ritalin). Journal of Child Psychology and Psychiatry, 1971, 12, 129-139.

Weiss, G.: The natural history of hyperactivity in childhood and treatment with stimulant medication at different ages: a summary of research findings. International Journal of Mental Health, 1975, 4, 213-226.

Wender, P.: Minimal Brain Dysfunction in Children, New York, Wiley, 1971.

Werry, J.: Developmental hyperactivity. Pediatric Clinics of North America, 1968, 15, 581-599.

Werry, J.: Organic factors in chilhood psychopathology, in Quay, H. and Werry, J. (eds) Psychopathological Disorders of Childhood, New York, Wiley, 1972, 83-121.

Werry, J. and Aman, M.: Methylphenidate and haloperidol in children. Effect on attention, memory and activity. Archives of General Psychiatry, 1975, 32, 790-795.

Werry, J. and Quay, H.: Observing the behavior of elementary school children. Exceptional Children, 1969, 35, 461-469.

Werry, J. and Sprague, R.: Hyperactivity in Costello, C. (ed) Symptoms of Psychpathology, a Handbook. New York, Wiley, 1970, 397-416.

Werry, J. and Sprague, R.: Methylphenidate in hyperactive children: effect of dosage. Australia and New Zealand Journal of Psychiatry, 1974, 8, 9-19.

Winsberg, B., Bialer, I., Kupietz, S. and Tobias, J.: Effects of imipramine and dextroamphetamine on behavior of neuropsychiatrically impaired children. American Journal of Psychiatry, 1972, 128, 1425-1431.

Zimet, S.C., Camp B.W. and White K.: Instructions for the classroom observation form Appendix B. In B.W. Camp, a pilot study of verbal learning in young aggressive boys. Final report to the National Institute of Education on Grant No. NEG-00-3-0029, February, 1977.

APPENDIX V

PERFORMANCE TESTS FOR PEDIATRIC PSYCHOPHARMACOLOGY STUDIES*

INTRODUCTION

Definitions

Performance test is a term which will be used in this paper to mean commonly administered psychological tests which measure some aspect of a child's behavior and result in a quantitative, standardized score. Only those standardized tests which have empirical data on reliability and validity or have been used successfully in psychotropic drug studies will be discussed.

Performance tests should be contrasted with other measuring instruments which may also result in a numerical score but which are not based specifically upon elicited, observable behavior. Examples of other measuring devices are rating scales from which one can obtain numerical scores of the judgments of a child's caretaker, such as a teacher (see Conners' paper), and observational instruments from which one can obtain quantified information about the child's behavior in a given setting from a trained observer (see Werry's paper).

There are several advantages to using performance tests in pediatric psychopharmacology studies: (1) these measures tend to be more sensitive to drug effects because there is less error of measurement than with the other techniques; (2) the measures are generally straightforward and objective with specific instructions, thus the tests can readily be administered in many laboratories and the results, hopefully, replicated; and (3) these performance tests tend to be based in psychological theory, thus the results should be useful in anchoring current nontheroetical, pediatric psychopharmacology to basic theories of behavior.

A very limited number of review articles have been written about the effects of psychotropic drugs on performance tests. One of the first such articles was written by Hartlage (1965) who comprehensively reviewed the effects of chlorpromazine on learning and intelligence in both animals and adults. Wolfensberger and Menolascino (1968; 1970) also reviewed the literature and then the methodological issues in evaluating the effects of psychotropic drugs on the intellectual performance of the mentally retarded.

For reviews of performance tests used, as well as a coverage of other issues, refer to Sprague and Werry (1971),for mentally retarded children, for emotionally disturbed and autistic children Campbell (1973; 1975), for the hyperactive child (Sprague & Sleator, 1975; Sprague & Werry, 1974; Winchell, 1975), and for the enuretic child, perhaps the most common childhood disorder treated with psychotropic drugs, Blackwell and Currah (1973). A basic reference for all types of psychological and performance tests is Buros (1972).

In 1973, I wrote a brief review of performance tests in the context of recommended behavioral measures for pediatric psychopharmacology studies as part of a special issue, Pharmacotherapy of Children, of the Psychopharmacology Bulletin (Sprague, 1973). The Psychopharmacology Research Branch of the National Institute of Mental Health has been actively involved in psychotropic studies for a number of years. Early in the work with adult patients it became apparent that some type of standardized battery of measures would be highly useful in comparing results of studies from one laboratory to another. Thus, the

*Written by Robert L. Sprague

ECDEU (Early Clinical Drug Evaluation Unit) was founded as a part of the Psychopharmacology Research Branch. The unit has published standardized tests and instruments, distributed forms for these measures to investigators, and assisted in the statistical analysis of experiments as a way of developing a central depository of information about psychotropic drugs.

Shortly after the special issue was published, Knights (1974) prepared a very interesting paper in which he surveyed 18 psychotropic drug studies with children which had used psychological tests to monitor behavioral changes. He attempted to empirically assess the sensitivity of the tests. The 18 drug studies were limited to studies which met the following criteria:

(1) used children as subjects with learning problems, MBD, or hyperactivity,

(2) used a double-blind design with placebo control including pre- and post-psychometric assessments;

(3) presented a statistical analysis of the test, and

(4) administered one or more of three drugs:

1. dextroamphetamine (8 studies),

2. methylphenidate (8 studies),

3. pemoline (2 studies).

As an index of sensitivity of a test to measure drug effects, he obtained a percentage significance score by dividing the number of times a test was statistically significant by the number of times it was given in the studies. He reported that a total of 49 different tests were administered for an average of 5.3 tests per study. A wide range of percentages for the index was obtained from a high of 66% significant with the Porteus Mazes to a low of 5% for all 10 subtests of the WISC.

Using what he termed a "rational approach" to test classification, Knights attempted to group the tests according to the basic psychological process being tapped by the test. He listed 11 categories:

(1) motility

(2) complex motor

(3) attention and vigilence

(4) new learning

(5) intelligence

(6) visual-motor and spatial

(7) auditory perception and memory

(8) verbal fluency

(9) simple motor

(10) language and achievement

(11) problem solving

Again, calculating a sensitivity index category by category, the range varied from a high of 40% for motility to the low of 2% for problem solving.

More recently Kleinknecht and Donaldson (1975) have reviewed 23 studies which were conducted to assess the cognitive and psychomotor effects of diazepam on adults. As is also true of the pediatric psychopharmacology literature, these authors point out that the vast bulk of the studies which focused on the cognitive effects have been published since 1970--20 of the 23 studies. What is important in this context, however, is the groups developed by the authors to categorize the 40 to 50 different tests reported in their survey:

(1) reflex speed

(2) critical flicker fushion threshold

(3) decision making

(4) learning and memory

(5) concentration and vigilence

(6) perceptual motor performance

CATEGORIES OF PERFORMANCE TESTS

Combining the grouping by basic psychological process of Knights and the empirical classification of Kleinknecht and Donaldson, I am suggesting six categories of pediatric performance tests which reflect both practical and research considerations. These categories are:

(1) intelligence

(2) achievement

(3) motor and motility

(4) learning and attention

(5) visual motor

(6) auditory and verbal

1. Intelligence Tests

Intelligence tests tap a wide variety of basic psychological processes and are not in any sense simple measures of single psychological functions. But intelligence tests have become standard parts of pediatric assessments, and for this reason are listed separately here. Almost 30 years ago Wechsler (1949) published an intelligence test for children which differed greatly from the traditional Stanford-Binet test that had been used almost exclusively with children prior to that time. The Wechsler Intelligence Scale for Children (much more commonly known by the acronym WISC) was subdivided into a verbal and performance parts with a total of six subtests in each part and a quantitative score derivable for each subtest. The norms extended from 5 years to 16 years in age.

In a long and productive series of studies, Conners (1973) has repeatedly used the WISC and reported that many of its subtests are sensitive to psychotropic drug manipulations.

Wechsler (1974) updated the WISC with a revision, WISC-R. The range of the norms has been changed, it is from age 6 to 17 years. The basic format of verbal and performance parts with six subtests each for a total of 12 remains the same. Another version of the

WISC, WPPSI (Wechsler Preschool and Primary Scale of Intelligence), which extended the norms downward to 4 years of age was published by Wechsler (1963). Otherwise, the WPPSI followed the basic format of the original WISC.[1]

The Porteus Maze Tests were published by S.D. Porteus more than sixty years ago (Porteus, 1915)[2]. After fifty years of usage, the author published another book on the test (Porteus, 1965). Then in the mid 1960's, largely through the work of Conners, the Porteus Maze Test was popularized as a measure sensitive to psychotropic drugs (Conners, 1972a; 1972b; 1973).

The tests consist of twelve mazes graded in difficulty from year III to adult I. The examinee is required to trace a path with a pencil from the start to the end without touching or crossing a boundary line and without entering a deadend. The task apparently requires the subject to look ahead and plan carefully his movement. Impulsive action results in many errors.

2. Achievement Tests

Although there are numerous tests commercially available designed to assess the amount of learning gained from classroom instruction, most of the tests are lengthy and/or are designed for group administration to the class as a whole by the teacher. One of the few achievement tests which is short and easy to give is the Wide Range Achievement Test (Jastak & Jastak, 1965)[3]. The test provides oral reading, spelling, and arithmetic computation scores from kindergarten through college levels.

3. Motor and Motility Tests

Motor Development

In a recent article, Lewko (in press) surveyed 400 facilities serving exceptional children to ascertain what tests of motor ability were being given on a widespread basis. Although more than 250 tests were reported, most of them unpublished, only four were given widely:

(1) The Denver Developmental Screening Test

(2) Gesell Developmental Schedules

(3) Purdue Perceptual-Motor Survey

(4) Lincoln-Oseretsky Motor Development Scale

The Lincoln-Oseretsky Test was standardized by Sloan (1955) on a population of children from central Illinois after being obtained from the items developed by a Russian, N. Oseretsky. Since it is time consuming and probably boring for both examiner and child, it is not recommended for routine use.

[1]The WISC can be purchased for $29.50, the WISC-R for $35.00, and the WPPSI for $29.00 from the Psychological Corporation, 304 East 45th Street, New York City 10017.

[2]The Porteus Maze Test can be purchased for $16.50 from the Psychological Corporation, 304 East 45th Street, New York City 10017.

[3]The Wide Range Achievement Test can be purchased for $5.20 from the Psychological Corporation, 304 East 45th Street, New York City 10017.

Note: Footnotes list suggested vendors for the commercially available tests mentioned in the text. The suggested prices may have increased by the time of publication of this guideline.

The Gesell Developmental Schedules (Gesell & Thompson, 1938) is a series of schedules of motor tasks for infants and young children from the age of 4 weeks to 72 months which has been standardized.[4]

The Denver Developmental Screening Test (Frankenburg, Dodds, & Fandal, 1970) was standardized on children from 1 month to 6 years of age. It is divided into four parts, only two of pertinence in this context: Fine Motor-Adaptive and Gross Motor.

Roach and Kephart (1966) developed The Purdue Perceptual-Motor Survey for assessing the motor ability of children from 6 to 10 years of age.

Motor Steadiness Test

To assess the behavioral effects of brain damage, Reitan (1966) has been developing a series of behavioral tests. Parts of these tests have been extended downward in age for the use with children (Reitan, 1973). Of particular interest is the Motor Steadiness Battery described by Klove (1963) and standardized by Knights and Moule (1968). The test consists of a finger maze which the child traces with an electrical stylus (any contact with the side counts as an error), a graduated series of holes in which the child holds a stylus in the hole without touching the sides, and a peg board test.

If one is interested in measuring the fidgeting which occurs when a child is seated in a chair performing a sedentary task, the stabilmetric cushion described by Sprague and Toppe (1966) has been shown to be useful and sensitive to psychotropic drug manipulations (Sprague & Sleator, 1973). There are some disadvantages to the stabilmetric cushion, the primary one being that it can only be utilized in settings where the child is expected to be seated.

Playroom Measures

Many facilities, such as child guidance clinics, often have some kind of playroom for children. With a moderate amount of effort, a playroom can be standardized with toys so that a child can be observed in the room as he plays with various toys and moves about. Routh et al. (1974) has been the latest of several investigators (Hutt, Hutt, & Ounsted, 1963) to describe measures which can be obtained from a standardized playroom.

4. Learning and Attention

Matching Familiar Figures

Except for the Matching Familiar Figures Test, the other tests listed in this section require instrumentation and have, thus, been listed last. The Matching Familiar Test is a task that requires the child to look at simple, familiar line drawings and then select from several, similar figures the one identical to the original. It has been used numerous times with hyperactive children and is sensitive to psychotropic drug effects (Douglas, 1972).

Continuous Performance Test

Conners and Rothschild (1968) described the Continuous Performance Test which is a task that requires continuous monitoring and vigilence on the part of the child to detect an infrequent target stimulus among other stimuli that are repeatedly presented at a fast rate of speed, i.e., one every 1.5 to 2 seconds. Distractable and hyperactive children soon tire of the task and make more errors than normal children. To utilize the test, the investigator needs projection equipment and equipment that can measure latencies of

[4] The Gesell Developmental Schedules can be purchased for $132.00 from the Psychological Corporation, 304 East 45th Street, New York City 10017

responding to at least 0.1 of a second. However, if funds are available for purchase of equipment, it is a useful, sensitive test of psychotropic drug effects.[5]

Picture Recognition Task

Another task which requires projection equipment and timing equipment as well as some kind of automatic device to either print or punch out data, is the picture recognition task of short-term memory. The task consists of presenting to the child a series of arrays of pictures (ranging in size from 1 to 15 pictures), allowing a few seconds to pass after presentation of the picture, and then turning on two test pictures and requesting the child to indicate which one he had seen in the previously presented stimulus array. The task was first described by Scott (1971) who used it extensively. In psychotropic drug studies it has proven sensitive to drug manipulations and, perhaps more important, dosage manipulations (Sprague & Sleator, 1975).[6]

5. Visual Motor

Bender-Gestalt Test

A number of commercial tests are available which require the child to integrate visual perception with movement (visual motor), e.g. look at a diagram and then reproduce it by drawing it. But only one of these tests has been used a number of times as a measure in psychotropic drug studies (Conners, 1967, 1973). The Bender-Gestalt (Bender, 1946) is a test consisting of eight cards with a series of drawings on them. The card is shown to the child for a few seconds, then the child is requested to reproduce the drawing as well as he can from memory. A standardized scoring system (Kopitz, 1964) can be used to obtain quantitative information from the test.[7]

6. Auditory and Verbal

Illinois Test of Psycholinguistic Ability (ITPA)

Again, as has been mentioned before, there are a number of commercially available tests which tap either the auditory perception of the child or the verbal productions of the child, but most of these tests have not been used in drug studies, consequently there is no evidence to indicate whether the tests might be sensitive to drug effects. The exception to this is the ITPA (Kirk & McCarthy, 1961; Kirk, McCarthy & Kirk, 1968). The ITPA consists of a series of twelve subtasks requiring a variety of auditory and visual decoding and encoding skills. The test has been recommended by Conners (1967, 1972) for use in drug studies.[8]

[5]Companies making equipment like this are: BRS, 5301 Holland Drive, Beltsville, Maryland 20705; Grason-Stadler, Concord, Massachusetts 01742; Lafayette Instrument Co., P.O. Box 1279, Lafayette, Indiana 47902.

[6]For more details about equipment, contact Behavioral Apparatus Builders, P.O. Box 775, St. Joseph, Illinois 61873.

[7]The Bender-Gestalt can be purchased for $11.75, and the Koppitz text for $8.75, from the Psychological Corporation, 304 East 45th Street, New York City 10017.

[8]The ITPA can be purchased for $57.50 from Western Psychological Services, 12031 Wilshire Boulevard, Los Angeles, California 90025.

Michigan Word Naming Test

Verbal fluency in word naming has not been studied extensively, but in one experiment (Creager & van Riper, 1968) significant differences were reported between placebo and methylphenidate on the number of words named using the Michigan Word Naming Test (Morpurgo, 1953). This test is mentioned here because it is thought that verbal fluency is an important area of children's school performance and should be further investigated.

CONCLUSIONS:

The basic conclusion that one can draw after reviewing a series of psychotropic drug studies with children is that there are very few standardized psychometric instruments that have been used sufficiently to be recommended as sensitive, reliable measures to detect drug and/or dosage differences. It is recommended that an investigator who plans a psychotropic drug study with children include one of the standard intelligence tests, achievement tests, and one of the measures of learning and attention as a minimum battery. Then other tests or experimental tasks may be added at the experimenter's predilection. As of now, there simply is not enough data to recommend a more extensive battery than these three categories.

REFERENCES

Bender, L. Instructions for use of visual motor Gestalt test. American Orthopsychiatric Association, 1946.

Blackwell, B., & Currah, J. The psychopharmacology of nocturnal enuresis. In I. Kalvin, R.C. MacKeith, and S.R. Meadow (Eds.), Bladder control and enuresis. London: William Heinemann Medical Book, 1973.

Buros, O.K. The mental measurements yearbook (7th ed.). Highland Park, N.J.: Gryphon Press, 1972.

Campbell, M. Biological interventions in psychoses of childhood. Journal of Autism and Childhood Schizophrenia, 1973, 3, 347-373.

Campbell, M. Psychopharmacology in childhood psychosis. International Journal of Mental Health, 1975, 4, 238-254.

Conners, C.K. The syndrome of minimal brain dysfuction: Psychological aspects. The Pediatric Clinics of North America, 1967, 14, 749-766.

Conners, C.K. Pharmacotherapy of psychopathology in children. In H.C. Quary & J.S. Werry (Eds.), Psychopathological disorders of childhood. New York: Wiley, 1972. (a)

Conners, C.K. Symposium: Behavior modification by drugs II. Psychological effects of stimulant drugs in children with minimal brain dysfunction. Pediatrics, 1972, 49, 702-708. (b)

Conners, C.K. Psychological assessment of children with minimal brain dysfunction. Annals of the New York Academy of Sciences, 1973, 205, 283-302.

Conners, C.K., & Rothschild, G.H. Drugs and Learning in children. In J. Hellmuth (Ed.), Learning Disorders (Vol. 3), Seattle: Special Child Publications, 1968.

Creager, R.O., & van Riper, C. The effect of methylphenidate on the verbal productivity of children with cerebral dysfunction. Journal of Speech and Hearing Research, 1967, 10, 623-628.

Douglas, V.I. Stop, Look, and Listen: The problem of sustained attention and impulse control in hyperactive and normal children. Canadian Journal of Behavior Science, 1972, 4, 259-282.

Frankenburg, W.K., Dodds, J.B., & Fandal, A.W. Denver developmental screening test manual. Denver: University of Colorado Medical Center, 1970.

Gesell, A., & Thompson, H. The psychology of early growth. New York: Macmillan, 1938.

Hartlage, L.C. Effects of chlorpromazine on learning. Psychological Bulletin, 1966, 64, 235-245.

Hutt, C., Hutt, S.J., & Ounsted, C. A method for the study of children's behavior. Developmental Medicine and Child Neurology, 1963, 233-245.

Jastak, J.F., & Jastak, S.R. The wide range achievement test. Wilmington, Deleware: Guidance Associates, 1965.

Kirk, S.A., & McCarthy, J.J. The Illinois Test of Psycholinguistic Abilities—An approach to differential diagnosis. American Journal of Mental Deficiency, 1961, 66, 399-412.

Kirk, S.A., McCarthy, J.J., & Kirk, W. Examiner's manual Illinois test of psycholinguistic abilities. Urbana, Illinois: University of Illinois, 1968.

Kleinknecht, R.A., & Donaldson, D. A review of the effects of diazepam on cognitive and psychomotor performance. The Journal of Nervous and Mental Disease, 1975, 161, 399-411.

Klove, H. Clinical neuropsychology. In F.M. Forster (Ed.), The medical clinics of North America. New York: Saunders, 1963. Pp. 1647-1658.

Knights, R.M. Psychometric assessment of stimulant-induced behavior change. In C.K. Conners (Ed.), Clinical use of stimulant drugs in children. The Hague: Excerpta Medica, 1974.

Knights, R.M., & Moule, A.D. Normative data on the motor steadiness battery for children. Perceptual and Motor Skills. 1968, 26, 643-650.

Koppitz, E. The Bender-Gestalt Test for young children. New York: Grune & Stratton, 1964.

Lewko, J.H. A survey of current practices in evaluating motor behavior of disabled children. The American Journal of Occupational Therapy, in press.

Morpurgo, C.V. The Michigan Picture Test. Michigan Department of Mental Health, Science Research Association, Inc., 1953.

Porteus, S.D. Mental tests for feeble-minded: A new series. Journal of Psycho-Asthenics, 1915, 19, 200-213.

Porteus, S.D. Porteus maze test: Fifty years' application. Palo Alto, California: Pacific Books, 1965.

Reitan, R.M. A research program on the psychological effects of brain lesions in human beings. In N.R. Ellis (Ed.), International review of research in mental retardation (Vol. 1). New York: Academic Press, 1966.

Reitan, R.M., & Boll, T.J. Neuropsychological correlates of minimal brain dysfunction. Annals of the New York Academy of Sciences, 1973, 205, 65-88.

Roach, E.C., & Kephart, N.C. The Purdue perceptual-motor survey. Columbus, Ohio: Merrill, 1966.

Routh, D.K., Schroeder, C.S., & O'Tuama, L.A. Development of activity level in children. Developmental Psychology, 1974, 10, 163-168.

Scott, K.G. Recognition memory: A research strategy and a summary of initial findings. In N.R. Ellis (Ed.), International review of research in mental retardation (Vol. 5). New York: Academic Press, 1971.

Sloan, W. The Lincoln-Oseretsky Motor Development Scale. Genetic Psychology Monographs, 1955, 51, 183-252.

Sprague, R.L. Recommended performance measures for psychotropic drug investigations. Pharmacotherapy of Children. Special Issue of Psychopharmacology Bulletin, 1973, 85-88.

Sprague, R.L., & Sleator, E.K. Effects of psychopharmacologic agents on learning disorders. Pediatric Clinics of North America, 1973, 20, 719-735.

Sprague, R.L., & Sleator, E.K. What is the proper dose of stimulant drugs in children. International Journal of Mental Health, 1975, 4, 75-104.

Sprague, R.L., & Toppe, L.K. Relationship between activity level and delay of reinforcement. Journal of Experimental Child Psychology, 1966, 3, 390-397.

Sprague, R.L., & Werry, J.S. Methodology of psychopharmacological studies with the retarded. In N.R. Ellis (Ed.), International review of research in mental retardation (Vol. 5). New York: Academic Press, 1971.

Sprague, R.L., & Werry, J.S. Psychotropic drugs and handicapped children. In L. Mann & D.A. Sabatino (Eds.). The second review of special education. Philadelphia: JSE Press, 1974.

Wechsler, D. Wechsler Intelligence Scale for Children. New York: Psychological Corporation, 1949.

Wechsler, D. Manual for the Wechsler preschool and primary scale of intelligence. New York: Psychological Corp., 1963.

Wechsler, D. Manual for the Wechsler intelligence scale for children--revised. New York: Psychological Corp., 1974.

Winchell, C.A. The hyperkinetic child: A bibliography of medical, educational, and behavioral studies. Westport, Conn.: Greenwood Press, 1975.

Wolfensberger, W., & Menolascino, F. Basic considerations in evaluating ability of drugs to stimulate cognitive development in retardates. American Journal of Mental Deficiency, 1968, 73, 414-423.

Wolfensberger, W., & Menolascino, F.J. Methodological considerations in evaluating the intelligence-enhancing properties of drugs. In F.J. Menolascino (Ed.), Psychiatric approaches to mental retardation. New York: Basic Books, 1970.

APPENDIX VI

GLOBAL RATING SCALES FOR CHILDHOOD PSYCHOPHARMACOLOGY*

This discussion will deal with so-called "global" rating scales for children which may be appropriate for use as selection and dependent measures in drug studies. These scales or observation schedules rely on the observer to synthesize primary observational data into judgements, adjectives, descriptions or classes rather than to directly observe, count, record or characterize ongoiong at the time at which it occurs.

The distinction between a direct observation and a global judgment is not absolute: even direct time-sampling methods require some degree of integrative judgment by assigning a carefully defined behavior to some class (e.g., "hitting", or "on-task"). Virtually no behavior can be said to occur in the absence of coding rules used by the observer. It is frequently assumed that a direct observation (such as time-sampling or interval sampling) is more accurate than a judgment made after the fact in which a number of behaviors are subsumed under one trait name by an observer. This is in fact a very knotty problem going to the heart of measurement and epistemology in behavioral science.

One of the major issues has to do with the relevance or meaningfulness of behaviors selected for direct observation. This issue was one of the key problems addressed in Murray's seminal Explorations in Personality (1938). Murray noted that:

> Some psychologists may prefer to limit themselves to the study of one kind of episode. For instance, they may study the responses of a great number of individuals to a specific situation. They may attempt to discover what changes in the situation bring about important changes in response. But, since every response is partially determined by the after-effects of previous experiences, the psychologist will never fully understand an episode if he abstracts it from ontogeny, the developmental history of the individual. (p. 2)

He goes on to distinguish between short motor units of behavior, which he called "actones", and units related to some adaptive goal of the organism and its environment ("thema"). This point of view eschews limited time samples of behavior because of the difficulty of relating the behavior to a meaningful pattern of which the organism is a part. Without wishing to revive the many arguments psychologists have inflicted upon each other since first regarding this issue, we may simply note that this issue of the behavior-in-isolation, vs. the behavior in context of previous history and environment, is still one that confronts every observer who wishes to abstract from the total flow of behavior those elements that are useful. What is useful is usually considered to be that which is reliable (repeatable, agreed upon by others) and which is also valid (measures what is says, predicts); and, one might add, relevant to the purposes of the study at hand.

The process whereby an observer comes to abstract and synthesize some specific samples of behavior is in fact a somewhat mysterious and unknown matter. It is known, however, that the 'set' of the observer, his degree of contact with the subject, his language framework, his values - are variables influencing the act of observing. In the simplest sense, the ego of the observer influences both what he chooses to look at and how he characterizes what he sees. Moreover, the "Heisenberg Principle" of observing applies: most observers are in the process of shaping the behavior they are ostensibly recording or observing, a fact often salient in the parent and teacher observing a child, which is one basis for claiming that such observers have limited value in an 'objective' measurement of behavior.

*C. Keith Conners

One of the important landmarks in the measurement of behavior is Osgood's Measurement of Meaning (1957). In his studies Osgood found that an enormous number of adjectives applied to behavior could frequently be reduced to three sources of variance: an evaluative dimension (good-bad), a power dimension (strong-weak), and an activity dimension (fast-slow). This formed the basis for his widely used technique of the semantic differential. The point of those studies in this context is that these categories appear to strongly influence almost all human judgments involving the assignment of a meaning to ongoing samples of behavior; like Kant's use of concepts such as space-time, they appear to form the windows through which reality is observed. Thus, a teacher looking at a child in the classroom uses a strong evaluative dimension: she relates the behavior to the rules, structure and purposes of the classroom, to her own concept of good-bad in the classroom context (and perhaps elsewhere as well). Parents who judge a child "active" must perforce do so with respect to their own internal standards of what is permissible or desirable, not just to "what is" or "out there". Behaviorists typically try to minimize factors such as an evaluative frame of reference by carefully defining the rules for classifying a particular motor act, but this is often an ideal rather than a fact. A particular vocalization or motor act will still frequently be characterizied by reference to its meaning in a social context (what constitutes "hitting" or "swearing" can be remarkably vague and requires long lists of qualifiers to achieve reliability). The questions one has after such a definition is arrived at, and the behavior appropriately sampled, is "so what?" "Does it matter, does it relate to anything in the real world, is it sensitive to environmental manipulation?" - These are empirical questions which may give disappointing answers. For example, Werry has commented in his chapter that such direct observational measures can be disappointingly insensitive to drug effects easily detected by global ratings. In a recent study, Roberta Ray and colleagues at the University of Wisconsin found no effects whatsoever in an experiment when classroom, time-sampling observations were used, but they found what appeared to be real effects on the teacher's global ratings. One explanation for these findings may have to do with ability of the observer to evaluate the behavior in a context with respect to standards or meanings supplied by the observer. Since these are minimized in the direct observation samples, the latter may be less sensitive to certain interventions; the tiny bits of behavior may be too tiny to catch the relevance of a larger pattern or whole which the observer supplies. Some might argue that this approach leads to a naive subjectivism and away from what is "in the child" or "out there". But again, this appears to be a matter for empirical study, and to date, global rating scales often come out much better in terms of utility, sensitivity to environmental changes, and ability to predict other classes of behavior.

The fact seems to be that qualities of a child such as "impulsive" or distractible" are not directly observable but require a sufficiently long sample of behavior in order that these adjectives can be applied by someone who has certain standards of these concepts. "Restlessness" may show very little correlation with actual wiggling, running, fidgeting, etc. because this is a category that observers employ when a certain quality of behavior reaches some threshold, probably a threshold related to the tolerance of the observer rather than to some intrinsic property of the child; or it is a quality which is inferred or attributed to the child by matching a sample of behavior against some internal schema of how goal-directed the behavior is rather than its actual quantity. One might argue that such scales are in fact the most meaningful from a behavioral point of view precisely because they abstract behaviors that have social significance. A rat in an activity cage may produce so many counts of activity, but whether such counts are relevant to another rat in the cage may depend on factors quite different than those relevant to the psychologist looking at the counter.

REVIEW OF SCALES PRIOR TO 1973*

The purpose of this report is to describe some rating scales for use in children's drug studies. It seems eminently clear that no single choice of scales is likely to meet the needs for the

*Reproduced from ECDEU Assessment Manual for Psychopharmacology: Revised, 1976. William Guy (ed.) DHEW Publication No. (ADM) 76-338.

variety of populations, designs, facilities and purposes of various research problems, and though I have chosen to recommend certain scales for consideration, I have also presented alternatives that may enrich the discussion and possibly be of use to investigators unfamiliar with these alternatives.

A number of good sources are available regarding the technology of scale construction and methodologic issues (1, 2, 3), and reviews of rating scales in psychiatric settings are available (4, 5). While there is indeed an elaborate technology for producing "pure" psychometric instruments, most evidence seems to indicate that the practical gains from elaborate and sophisticated scaling procedures is minimal (1), and I do not propose to deal with the many methodologic issues raised in the use and construction of rating scales. Certain basic attributes of reliability and validity need, of course, to be considered, and for the most part I have not included a number of scales that look interesting but which have no published reliability or validity data.

The choice of children's rating scales needs to be based on certain criteria and working assumptions which will eliminate some scales from further consideration.

First, there is the source of the rating data. If the source of data is the parent or teacher, then the scale must be non-technical, brief and easily filled out. A clinician or trained observer on the other hand, may use much more detailed and theoretically-oriented instruments. Since parent, teacher, and clinician have different (though overlapping) behavior samples, the scales for different observers almost certainly need to be different in content, though an overlap in some areas would be desirable.

Secondly, there is the question of level of observation. This can be very molecular--where specific behavioral acts or sequences can be observed and time-sampled--or the categories can be quite global, abstract or inferential. Most people are agreed that ratings which require a great deal of inference about underlying processes tend to be unreliable; but descriptive global ratings that use "middle level" inferences are often most reliable. Unless the observer is highly trained there is likely to be a loss of reliability for rating of molecular events. We have, therefore, tended to assume that some middle level of abstraction, requiring a minimum of inference, is preferable unless highly trained observers are available.

A related issue is whether one is interested in rating current behaviors, symptoms or states; or whether the intent is to describe basic traits, dispositions, or personality characteristics. While not mutually exclusive, these approaches lead to somewhat different types of scales. I have assumed that a symptom focus is most appropriate for our purposes, though the difference between a symptom and a trait is probably more a question of values as to whether the behavior in question is normative or undesirable.

Whether one uses state or trait methods depends to some extent on the purpose of using the ratings in the first place. A use for prediction might well require more trait-disposition items while symptoms would seem to be more appropriate for measuring change. Both types of items are appropriate for questions of taxonomic classification. It is conceivable to me that all three purposes--predication, measurement of change, and classification--might be meaningfully applied in drug studies. In general, I have recommended the use of behavior items that are susceptible to short term change, but which can also be used in conjunction with statistical techniques for prediction and classification.

The population under study clearly makes a difference in the type of scale to be employed. It has seemed reasonable that separate instruments should be employed for severe psychiatric disturbances (psychosis, retardation, autism, etc.) as contrasted with the more frequent and typical patients found in out-patient settings. Institutionalized children are usually more severely affected by their illness, and many of their symptoms are of low frequency in outpatients (e.g., hallucinations, autistic alloofness).

Finally, the format of the scale needs consideration. For most purposes a scale with specific anchor points describing the behavior in question is most likely to be reliable and valid. But

such scales are also more cumberson and time-consuming to use. If the range of behavior to be sampled is broad, (as it is likely to be in the screening phase of a study) then the items should be brief and the rating procedure as simple as possible. This consideration has led me to recommend the "check-list" type of scale, especially for parent ratings.

Teacher Rating Scales

1. Cattel and Coan (6) administered a 38-item trait list of bipolar items to teachers of 198 first and second grade pupils. This list was compiled to include the major "markers" from other personality research, as well as "useful indicators of personality disturbance." Many of the items are probably irrelevant for symptom-oriented studies (e.g., "aesthetically sensitive, aesthetically fastidious, vs. lacking in arthistic feeling"), but for those investigators interested in predicting drug effect from personality traits, this might be a useful scale. They identified some 15 factors by Cattell's methods (oblique rotations), but the reliability of factor scores is not given, and the non-independence of the factors probably makes them of little use as independent predictors in regression equations.

2. Peterson (7) used the referral problems of 427 cases at a guidance clinic to select the 58 most common symptoms. The list was given to teachers of 831 kindergarten through sixth grade pupils for ratings. Two major factors (conduct problem and personality problem) emerged with considerable consistency across the whole age range. Interrater reliabilities (for the Kg sample) were .77 and .75 for factor scores for the two factors. Quite similar factors have emerged in a number of studies by Quay and associates (8) for various populations, from sources as disparate as case history ratings, questionnaires, standard ratings, and by a variety of factor extraction methods.

   However, several questions can be raised about these results. The presence of only two (sometimes three) factors suggests that either the repertoire of times is so restricted as to guarantee a small number of independent factors or the method of analysis produces few factors. Secondly, the two factors appear to subsume some very disparate behaviors which intuitively seem distinct. Thirdly, many of the items, particularly conduct problem items, are essentially synonyms guaranteeing that a strong factor will emerge. Some of the items are symptomatic (e.g. fighting) while some are essentially trait names (e.g. nervousness, aloofness). Nevertheless, similar factors emerge in some form or other in many other studies, and it is probably safe to assume that there are at least two important dimensions, or causally independent factors, that could be extremely useful in basic classification, prediction, and possibly measurement of change in drug studies.

3. A comprehensive classroom behavior and personality instrument has been developed by Shaeffer and colleagues at the Laboratory of Psychology of NIMH. The items were selected from a theoretical model of child behavior, have been extensively analyzed for factor structure and reliability, and tested in the U. S. and Scandinavia. Specific classroom behaviors are organized into traits, and the traits are organized into factors and arranged in a "circumplex" model. Figure 1 (see next page) shows the conceptualization of the item-trait-factor derivation, and Figure 2 is an example of the ordering of traits on a circumplex.[1] The major difficulty with this instrument seems to be its length. The 320 items in the scale seem prohibitively time-consuming for volunteer reporting by teachers. However, the excellent pool of items, and the extensive analytic work on sub-scales might be useful in some settings.

4. The Devereux Elementary School Behavior Rating Scale (9) is a 47-item anchored scale for teachers, with items easily grouped into 11 behavior factors. Normative data is available

[1] These data are from an unpublished manuscript by Shaeffer, Droppelmar, and Kaverboer. Unfortunately, at the time of this preparation I did not have available Dr. Shaeffer's most recent extensive work.

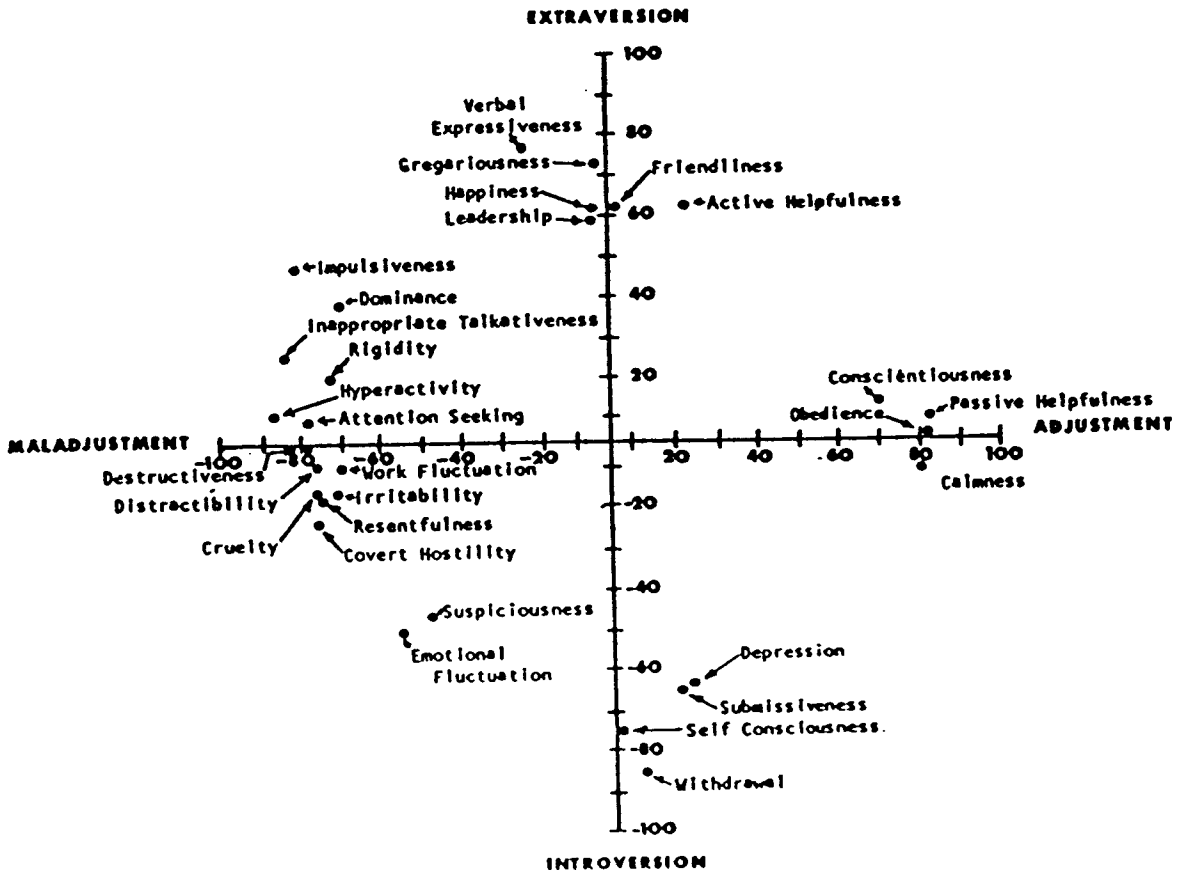
FIGURE 1

HIERARCHIAL STRUCTURE OF THE CLASSROOM BEHAVIOR INVENTORY
(Form for Preschool to Early Primary)

Molar, abstract → Concrete, specific

	Dimensions (3 Factors)	Traits (12 Scales)	Specific Behaviors (60 Items)
Classroom Adjustment	Extraversion vs. Introversion	Verbal Expressiveness Gregariousness	Always has something to say in group discussions.
		Social Withdrawal Self-Consciousness	Rarely joins in activities with others of his own accord.
	Love vs. Hostility	Considerateness Kindness	Is careful not to disturb an activity of another.
		Resentfulness Irritability	Sits and sulks if he has been reproved.
	Positive Task-Oriented vs. Negative Task-Oriented	Perseverance Concentration	Nearly always sticks to tasks until they are finished.
		Hyperactivity Distractibility	Frequently is twisting, turning or getting up from his chair.

FIGURE 2



on 809 normal children in kindergarten through 6th grades. Test-retest factor scale reliabilities range from .71 and .91 with small standard errors of measurement, and median reliability of .87. The factor structure is quite similar across grade levels. In general the scale meets most of the requirements for an instrument in drug studies, though I know of no demonstration that it is "drug-sensitive". This scale has a high priority for use as a standardized data- gathering instrument

5. A 39-item Teacher Symptom Checklist originally developed by Eisenberg and colleagues has been used in several drug studies and recently factor analyzed by Conners (10). The five factors are highly reliable on test-retest, and appear to be quite sensitive to changes due to drug, with relatively little placebo influence. Test-retest reliabilities over a one-month period ranged from .72 to .91. The five factors were labeled "aggressive conduct," "day-dreaming-inattentive," "anxious-fearful," "hyperactivity," "sociable-cooperative."

   A newer, slightly modified form has been developed which contains 10 items that overlap with the symptom checklist for parents, described below. This allows one to compare ratings from both sources on a common core of items.**

6. Two excellent teacher scales should be mentioned. Both are more appropriate for identification of learning disorders and children with developmental deviations than for measuring change, but in view of the likelihood of increased interest in drug studies of learning disorders, the scales are important to keep in mind where large scale screening may be needed to identify potential candidates for drug studies. The first is a 24-item anchored scale by Myklebust (11). The items are grouped into five areas: auditory comprehension and learning, spoken language, orientation (time, scale, relationship), behavior, and motor. The scale was used to identify children with minimal cerebral dysfunction in a sample of 2767 third and fourth graders. Excellent discriminative power and vailidity were shown with the scale, though reliabilities are not reported.

   The Classroom Screening Inventory developed by the Rocky Mountain Educational Laboratory (12) is an 80-item scale that is divided into 14 sub-scales focused on classroom learning and behavior. A very thorough item analysis, factor analysis, reliability and validity studies are reported. The instrument was used in a study of a stratified random sample of 2400 children in the Rocky Mountain area. Interrator reliability was .85. A validity study showed that the screening produced no false positives and very few false negatives. This instrument though still being developed is the best of its kind known to this writer.

   In summary, the Devereux Elementary School Behavior Rating Scale appears to meet most of the requisites for a brief, reliable scale for children's drug studies. As an alternative, the Conners scale is probably easier to use and less likely to be resisted by the busy teacher because of its checklist format. However, the more extensive published research on the Devereux Scale makes it appear as the best bet at this time.

Parent Rating Scales

A number of studies of the dimensions of symptom behavior in young children have been made during the past several years. Jenkins and Hewitt (13) described three clusters of traits identified from case records of 500 children rated on 90 symptoms. More recently, Jenkins (14) identified 5 clusters which he labeled "shy-seclusive," "overanxious-neurotic," "hyperactivity with poor concentration," "undomesticated," and "socialized delinquent."

**This 10-item scale is contained in the ECDEU manual from which the present material was taken.

These clusters fell into two broad categories of inhibited and aggressive children. Peterson (15) identified two dimensions from parent and teacher ratings which he labeled "conduct disorder" and "personality disorder." These patterns have emerged in several other studies by Quay (16), Dreger, et al. (17), and Borgatta and Fanshel (18). The latter study produced 12 factors: defiance, unsocialized tension-anxiety, lack of affection, infantilism, overcleanliness, sex precociousness, sex inhibition, learning difficulty, (a and b), likeability, responsibility. A second-order factor analysis produced six factors including an "acting-out" factor, developmental immaturity, inhibited behavior, learning disorder, and sociable-responsible. Reliabilities of factor scales are not given, but individual item reliability ranges from .60 to .77, suggesting that factor scales are likely to be highly reliable. These studies and others mentioned below provide a substantial base of knowledge for purposes of prediction and classification.

An anchored rating scale for nonprofessionals was developed by Spivack and Spotts (19) at the Devereux Foundation. Good norms are available for the 17 sub-scales of the 97-item scale. Like the teacher's version, this scale is throughly researched, easy to use and score, and covers a broad range range of psychopathology.

The Missouri Children's Bahavior Checklist (20) is a similar 70-item yes-no checklist of symptoms. The factors of aggression, inhibition, activity level, sleep disturbance, somatization and sociability have odd-even reliabilities ranging from .67 to .86. Inter-parent agreement on individual items ranged from 53% to 94%. Validity studies of clinic versus controls showed significant discrimination of all factors except somatization and sleep disturbance.

Conners (21) has described a 93-item parent symptom checklist that was factor-analyzed on 316 clinic patients between the ages of 6 and 14, and 367 normal controls of the same age. Twenty-four categories of symptoms (sleep, learning, sociability, etc.) were factor analyzed. Six factors were identified by principal components analysis and labeled aggressive conduct disorder, anxious-inhibited, anti-social, enuresis-encopresis, pychosomatic, and anxious-immature. Discriminant function analysis showed that 83% of controls and 70% of clinic patients could be correctly classified from factor scores. Neurotic and hyperkinetic children were also correctly identified in 77% and 74% of the cases, respectively. Mother-father agreement averaged .85 on total scores, but factor scale agreement is not reported as yet. The first two factors (conduct disorder and anxious-inhibited) have been used in drug studies and show significant drug-placebo interactions. A recently modified version employs a 10-item scale to overlap with teacher ratings for repeated measures in drug studies.

A factor analysis was also completed on individual items for the total sample of 683 subjects (previous analyses had shown close similarity in factor structure for different social classes, different age ranges, and for the sexes). Factor loadings on each of the seven factors are very similar to the factors reported by Achenbach, Borgatta and Fanshel (18), and several others.

One drawback of the scales described here is that none includes symptoms of severe psychopathology such as psychotic manifestations. A rather extensive study on children's psychiatric symptoms by Achenbach (22) includes more of such symptoms. The large, first principal component factor appeared to be a bipolar "internalizing vs. externalizing" factor, and the second large component was identified as a unipolar "diffuse psychopathology" factor. Eight rotated factors were identified as: somatic complaints, delinquent behavior, obsessions, compulsions and phobias; sexual problems; schizoid thinking, unsocialized aggression, hyperactivity; and one minor factor. The main problem with this scale is that it is designed for professionals or semi-professionals, so that various items would be difficult for parents to use (such as diplopia, compulsions, etc.). This is an excellent list, however, for rating of case reports or other symptom rating in a clinical context.

In summary, both the Conners and Devereux scales appear to be feasible in drug studies, with the latter scale being more thoroughly standardized.

Clinical Ratings

1. Very few standardized child-psychiatry rating scales are available. The brief standardized rating procedure described by Rutter and Graham (23) appears to have both good inter-examiner reliability and validity. A somewhat more comprehensive rating scale for psychiatrists has been provided by Drs. Klein from the Hillside Hospital but standardization procedures are not available at this time.

2. A valuable source of observation, particularly for measuring change in drug studies, is a behavior rating by the psychologist on the basis of observations made during psychological testing. I am unaware of any standardized forms for this purpose, but the rating scale used by the NINDS Collaborative Perinatal project appears to be excellent for most purposes.

Inpatients and Retarded

The Children's Behavior Inventory by Burdock and Hardesty (24) is a 139-item yes-no scale with items grouped by age-appropriateness. Extensive reliability and validity studies have been done, and the results indicate sufficient discriminative power and stability to warrant using the inventory in settings where a moderate amount of training of observers is possible. The items are rationally grouped into categories of vegetative function, appearance and mannerisms, speech and voice, emotional display, socialization and thought processes. Drug studies have not yet been reported with this instrument.

A much briefer scale has been reported by Davis, Sprague and Werry (25) for time-sampling measurement of sterotyped behavior in retardates. Interjudge reliabilities ranged from .61 to .88 for the 7 categories. The scale showed sensitivity to drug treatment, and would appear to be an excellent measure for this relatively restricted (but common) set of behaviors in retardates or other severely disturbed inpatients.

References for Review of Scales Prior to 1973

1. Chronbach, L. J. Essentials of Psychological Testing. Harper, 1960 (New York), 2nd Ed.

2. Lyerly, S.B. and Abbott, P. S, Handbook of Psychiatric Rating Scales (1959-1964). USPHS Publication #1495.

3. Guilford, J. P., Psychometric Methods, McGraw Hill, 1954 (New York), 2n Ed.

4. Norton, W. A. Review of Psychiatric Rating Scales, Canad. Psychiatric Assoc. Journal. 1967, 12 (6), 563-74.

5. Doty, D., Rating scales used in children's drug research: a review of the literature. In survey of research on psychopharmacology of children, by R. Sprague and J. Werry and Students (Children's Research Center, University of Illinois).

6. Cattell, R., and Coan, R. W., Child personality structure as revealed by teacher's behavior ratings. J. Clin. Psychol., 1957, 13, 315-327.

7. Peterson, D. R., Behavior problems in middle childhood J. Consult. Psychol., 1961, 25, 205-209.

8. Quay, H. C. Personality dimensions in deliquent males as inferred from the factor analysis of behavior ratings. J. Research in Crime and Deliquency, 1964, 1, 33-37.

9. Spivack, G. and Swift, M. The Devereux Elementary School Behavior Rating Scale. Devereux Foundation, Devon, Pennsylvania.

Spivak and Levine, M. The Devereux Child Behavior Rating Scales: a study of symptom behavior in latency age atypical children. Amer. J. Ment. Def., 1964,68, 700-717.

10. Conners, C. K. A teacher rating scale for use in drug studies with children. Amer. J. Psychiatry, 1969, 126, 884-888.

11. Myklebust, H., Minimal Brain Damage in Children. Final report to USPHS Contract #108-65-142.

12. Rocky Mountain Educational Laboratory Classroom Screening Instrument. RMEL, 1620 Reservoir R., Greeley, Colo. 80631.

13. Jenkins, R. L., and Hewitt, L., Types of personality encountered in child guidance clinics. Amer. J. Orthopsychiatry, 1944, 14, 8494.

14. Jenkins, A., Psychiatric syndromes in children and their relation to family background. Amer. J. Orthopsychiatry, 1966, 36, 450-457.

15. Peterson, D. R., op. cit.

16. Quay, H. C., op. cit.

17. Dreger, R. M., et al, Behavioral classification project. J. Consult. Psychol., 1964, 28, 1-13.

18. Borgatta, E.F. and Fanshel, D., Behavioral characteristics of children known to psychiatric outpatient clinics. Child Welfare League of America, 1965, (Library of Congress, 65-19746).

19. Spivack, G. and Spotts, the Devereux Child Behavior Rating Scale. Devereux Foundation, Devon, Pa.

20. Sines, J. O., Pauker, J. D., Sines, L. K. and Owen, D. R., Identification of clinically relevant dimensions of children's behavior, J. Consult. and Clin. Psychol., 1969, 33, 728-734.

21. Conners, C.K., Symptom patterns in hyperkinetic neurotic and normal children. Child. Devel. (in press).

22. Auchenback, T. M. The classification of children's psychiatric symptoms: a factor analytic study. Psychological Monographs, 1966, 80, No. 6.

23. Rutter, M. and Graham, P. The reliability and validity of a psychiatric interview for children. Brit. J. Psychiat., 1968, 114, 563-579.

24. Burdock, E. L., and Hardesty, Anne S., A children's behavior diagnostic inventory. Ann. N. Y. Acad. Sci., 1964, 105, 890-896.

25. Davis, K. V., Sprague, R. and Werry, J. Stereotyped behavior and activity level in severe retardates: The effect of drugs. Amer. J. Ment. Def., 1969, 73, 721-727.

26. Alderton, H. R. and Hoddinot, B. A., A controlled study of the use of thioridazine in the treatment of hyperactive and aggressive children in a children's psychiatric hospital. Canad. Psychiat. Assoc. J., 1964, 9, 120-130.

27. Pritchard, M. Observation of children in a psychiatric inpatient unit: design of a behavioral rating scale for nursing staff. Brit. J. Psychiat., 1963, 109, 572-578.

REVIEW OF RECENT RATING SCALES*

Bell, Waldrop and Weller (1972) described a rating system appropriate for nursery school age children in which characteristics of hyperactivty and withdrawal are rated.
As noted in our previous discussion of rating scales (1973), a conduct disturbance dimension and an anxiety-fearfulness dimension have emerged in most studies of diversity of traits in school age children. Bell et al's scale is an 11-point scale in which categories of frenetic play, induction of intervention, inability to delay, emotional aggression, nomadic play, and spilling-throwing are rated (hyperactivity dimension), and vacant staring, closeness to adult base, and chronic fearfulness are rated for the withdrawal factor. The scale items are anchored (e.g. at 1, 6, 9 and 11 on the scale). Results are based on observations of nursey school children at least two hours a day, with ratings made on a day-to-day basis or at the end of a month. Home visits were made to develop some of the ratings. The ratings were also summarized on a weekly basis in one of the studies. A factor scoring system based on normative data is provided from which one may compute a total hyperactivity and total withdrawal score and compare it with the optimal cutting point for differentiaiting the normal from extreme cases. Although comparisons with clinically diagnosed samples have not been made, and no drug studies carried out, this instrument should prove to be extremely useful in selecting subjects in the kindergarten or nursery school range for investigative studies. Reliabilities of the individual items is quite good, ranging from .59 to .94. The provision of a simple factor scoring system should also make the instrument useful for following children over time and detecting changes due to intervention. However, see Appendix VII for a more extensive review of preschool rating scales.

Blunden, Spring, and Greenberg (1974) carried out an extensive validation of their Classroom Behavior Inventory using 320 kindergarten boys. The scale uses ten categories of behavior associated with the hyperkinetic syndrome, with four individual items each rated on a 4 point scale ("not at all like the child"--1 point—to "very much like the child"—4 points.) A factor analysis showed that restlessness, impulsiveness, distractibility, low concentration and low perserverance loaded highly on factor one. Irritability and resentfulness loaded on factor three, while cheerfulness, social participation and verbal expression loaded on factor two. The fourth factor was uninterpretable.

Concurrent validity was measured by comparing the CBI with direct time-sampling in the classroom utilizing 15-second intervals over a 15-minute period three times for a week. Thus, each subject had 45 minutes of direct observation. Inter-observer agreement ranged from 71% to 78%, calculated by determining the ratio of the number of 15-second intervals in which the selected behavior was observed by both observers to the number of 15-second intervals in which the behavior was observed by at least one observer (This method is subject to spurious inflation as noted in Werry's chapter.)

The results were striking: only one of the CBI scales (impulsiveness) was actually significantly correlated with its direct observation counterpart (r=.50). Of the 49 correlations in the matrix, only nine were actually significant, with 6 of those being correlations of the direct observations with teacher's ratings of impulsiveness. Teachers made global judgements of whether the children had behavior disorders or not, and on 8 of 10 teacher ratings there were significant differences while only one of the direct observation scores differentiated the two groups (impulsiveness).

The authors suggest that either the low stability of the directly observed behaviors from the 45-minute sample, or limited inter-teacher reliabilities may have attenuated the correspondence of the two data sets. They also suggest that the teachers may have been essentially using only one "real" dimension, impulsiveness. However, one might equally well argue that the teachers' ratings were valid, and the directly observed behaviors invalid due to their highly context-specific, unrepresentive nature. Greenberg, et al (1972) have shown that the CBI is somewhat sensitive to drug effects, but once again we are left to wonder what is really being measured.

*This material was written for this appendix.

Davids (1971) has provided a clinical rating instrument for hyperkinesis which use 7 items rated on a 6 point scale. The instrument was published with full awareness that reliability and validity had not been established. It was used in a study comparing dexedrine and placebo by Denhoff, Davids, and Hawkins (1971). Three of the times (activity, short attention and impulsiveness) discriminated at a signficiant level between drug and placebo. Drug effects were prominent in those children whose teachers gave a rating of 4 or more on each of the six scales. Neither drug effects nor correlations with teacher scales were significant in the parent ratings using the same form. One recent study using the Davids instrument employed the Conners Teacher Rating scale as well, and the latter was caffeine sensitive while the former was not, but other than this unpublished study from the Montreal group (V. Douglas) I am unaware of comparisons between the two instruments.

The Children's Pathology Index (CPI) (Alderton and Hoddinot, 1968) is a scale for inpatient observation of children that has received careful study. This scale was not previously reviewed for consideration as a drug treatment measure but offers some value for inpatient settings. As described by the authors, the CPI

> ....consists of 38 categories each describing some type of disturbed function, behavior, attitude, relationship or emotional response. Each category contains five descriptive statements ordered from best adjustment (assigned a rank of 5) to worst adjustment (assigned a rank of 1). Each statement appears by itself, printed on a piece of paper 4-1/4" by 2". The five statements, and a numbered title sheet are stapled together to form a booklet. The 38 booklets together make up the instrument and are presented in numerical order...The rater's task is to select from each of the 38 booklets the statement that most typically describes the child.

The ranking of statements was verified by using six trained judges and computing coefficients of concordance, which ranged from 1.0 to .67, with 22 reaching 0.9 or better and all but five 0.8 or better. A factor analysis of the instrument produced four factors. The four factors, Disturbed Behavior Towards Adults, Neurotic Constriction, Destructive Behavior and Disturbed Self-perception appear to be similar to dimensions found on several other instruments; in particular the Conduct disorder and Anxiety factors seem to be constant dimensions of most instruments (see previous section).

Reliabilities were computed using four raters, 28 days apart and 42 days apart. The 28-day reliabilities were .85, .40, .79, and .79 for the four factors respectively; and the 42-day reliabilities were .75, .72, .79 and .88.

Concurrent validity was investigated by comparing time samples of aggessive behavior with the factor I scores. A correlation of .59 was obtained. A biserial correlation of .82 between factor I scores and psychiatrists' discharge prognosis for community adjustment. This means, of course, that the psychiatrist simply felt that more aggressive children would adjust more poorly. The actual patient status at 18 months after discharge was significantly associated with all 4 factors utilizing categories of institutionalized, remaining in the community and remaining in the community without significant difficulty. To what extent these findings reflect the self-fulfilling prophecy of the psychiatric discharge prognosis and recommendation is unclear. But it is notable that most of the effects are accounted for by the difference between the hospitalized and the non-hospitalized children—a result compatible with this hypothesis. If the psychiatrist both made a prognosis and assigned the children to other institutions, this would not reflect true independent predictive validity of the instrument. A further study showed that the CPI did not show significant inter-institutional profile differences in a comparison of 4 similar institutions. While this finding may imply the "universality" of the instrument as suggested by the authors, it could also be due to insensitivity.

The Deviant Behavior Inventory (DBI) is an instrument not found in general use, but intensively studied by Novick, et al (1966) and currently in use by this writer as a screening device for parents of children admitted to an inpatient unit. The value of the instrument appears to be in its careful wording of items (readability), its completeness (237 items), its

Q-sort administration, the use of a "not sure" category a clearly specified time reference, a procedure for self-correction of endorsements by partial re-sorting, and a focussed inquiry to document endorsed deviant behaviors. A careful look at the procedure of this study has much to recommend it for those who rely too cavalierly on parent-administered forms of this type. The authors comment that "It is apparent from our findings that despite all efforts to minimize the error due to false endorsements by reporters the residual error is of such magnitude as to seriously question the value of any behavioral assessment which does not take this into account." Specifically, they found that parents failed to pick as True a substantial number of items known from independent sources to be present; and conversely, that of those items picked as True, a substantial proportion were not ultimately judged to be invalid. Despite its limitations, this type of instrument serves a useful screening function by covering virtually all areas of symptomatology of relevance to the 8-12 year old age range, and if administered carefully can provide detailed parent descriptions useful in the evaluation of therapy or in follow-up. It is an instruement too long for frequent or repeated use, but the selection of target symptoms could be a useful way of generating an individualized scale for each patient of moderate length.

CONCLUSION

Very little new basic information on global rating scales has appeared. The scale of Blunden, Spring and Greenberg shows some promise for drug studies, but basic issues have not been resolved and are unlikely to be until careful comparison of different types of instruments are carried out. The reader should be aware of two major sourcebooks for reference use in the rating field: Comrey, Backer and Glaser (1973) have complied a source-book of over 1100 instruments; and Johnson and Bommarito (1971) have provided a review volume of tests and measures in child development. Several instruments in these compendia are of relevance to our review, but the scales described did not appear to have advantage over those mentioned, and none appear to have been tested in actual drug research.

References for Review of Recent Rating Scales

Alderton, H. Hoddinott, B: The children's pathology index. Canad. Psychiat. Ass. J. 13:353-361, 1968.

Bell R, Waldrop M, Weller G: A rating system for the assessment of hyperactive and withdrawn children in preschool samples. Amer. J. Orthopsychiatric. 42:23-34, 1972.

Bluden D, Spring C, Greenberg L: Validation of the classroom behavior inventory J. Consulting and Clinical Psychlogy 42:84-88, 1974.

Comrey A, Backer T, Glaser E: A sourcebook for mental health measures. Human Interaction Research Institute, 10889 Wilshire Blvd., Los Angeles, CA 90024. Prepared for NIMH, DHEW, 1973. 1100 instruments.

Davids A: An objective instrument for assessing hyperkinesis in children. J. Learn. Disabil. 4:35-37, 1971.

Denhoff Eric, Davids A, Hawkins A: Effects of dextroamphetamine on hyperkinetic children; a controlled double-blind study. J. Learn. Disabil. Nov., 1971.

Greenberg L, Personal Communication.

Johnson O G, Bommarito J W: Tests and measurements in child development: a handbook. San Francisco: Jossey-Bass, 1971.

Murray H, Exploration in Personality; A clinical and Experimental Study of 50 Men of College Age. New York, Oxford University Press, 1938, p 2.

Novick J, Rosenfeld E, Bloch D, et al: Ascertaining deviant behavior in children Journal of Consulting Psychology 30:230-238, 1966.

Osgood C, Suci G, Tannenbaum P: The Measurement of Meaning. Urbana, University of Illinois Press, 1957.

APPENDIX VII

REVIEW OF PRESCHOOL BEHAVIOR RATING SCALES

This chapter reviews behavior rating scales which are currently available for use in the preschool age period. It is limited to those which utilize parent or teacher reports of behavior rather than direct observations at home or school. The review does not approach exhaustive coverage of the field. Most of the scales reviewed have not been used in drug studies. Some, however, have been used in evaluating effects of other types of therapeutic interventions including the tincture of time. Those presented here were located through an ERIC search, review of Psychological Abstracts from 1972-1976, Index Medicus 1972, 1975 and through inclusion in Social-Emotional Measures for Preschool and Kindergarten Children (Walker, 1973), Tests and Measurements in Child Development: A Handbook (Johnson and Bommarito, 1971) and The Seventh Mental Measurements Yearbook (Buros, 1972). This process was supplemented by personal communication and search of recent publications in the Journal of Educational Psychology, Developmental Psychology, and Child Development.

Although the demand for psychotropic medication may be limited in the preschool age group, studies of any drugs to be administered on a chronic basis to young children should include evaluation of effects on behavior and psychological development.

The behavioral measures described here may be potentially useful in selecting patients for study, in documenting sample characteristics, and in evaluating drug safety and efficacy. Each of these uses may place different demands on the measuring instrument. The particular type of scale selected should be suitable for the use to which it will be put. When behavioral measures are used to determine whether a patient meets criteria for inclusion in a study, the measurement should have a high degree of reliability with appropriate normative or other background data to insure confidence that the patient population under study indeed meets the criteria defining it. Where behavioral measures are expected to change as a result of some drug effect, the sensitivity of the measure may be more important than its reliability. When a behavioral measure is to be used both as a selection tool and in measuring sensitivity to drug effects, the problem is more complex. In these instances, the more sensitive instrument is often selected while instability in the measuring instrument is accounted for by randomly assigning patients to receive active drug or placebo.

In general, rating scales which require the rater to determine whether a particular behavior is present or absent show greater inter-observer reliability and test-retest reliability than those which require some judgment regarding the degree to which a behavior is present. Similarly, those which provide a two-factor solution tend to be more reliable than those which yield a larger number of factors (see Behar for a discussion of this issue). Unless a large change in behavior is expected, the usefulness of scales using a "yes-no" format is most likely to be limited to establishing criteria for selection into the study, to describing the population and possibly to evaluating side-effects. Because of the great instability of even clearly deviant behavior in preschool children (Chamberlin, 1974; Schleifer and Weiss, 1974) these scales should not be used as the sole criteria for selecting children into a drug study.

Several summaries of pr school rating scales indicate that there are great similarities from scale to scale in the clusters of behaviors whicl they identify (Behar, 1974; Kohn and Rosman, 1972). Items in one scale have often been adapted from another scale and revised in a third. Two specific issues which appear repeatedly in studies of behavior ratings in preschool children have to do with the question of whether aggressive and hyperactive behavior should be grouped together into one dimension and the question of whether a particular behavioral dimension is unipolar or bipolar.

The extent to which aggressive-hostile behavior and hyperactive distractible behavior have been separated into different categories has depended upon the type of factor analysis employed and the item pool. Schaefer's (1971) discussion of the value of a three dimensional model points up the reason for the difficulty. If one seeks a two dimensional solution to the behavior domain, hyperactive-distractible behaviors cluster with aggressive-defiant behaviors at the maladjusted pole of a maladjusted-adjusted scale. When one employs a three dimensional model, hyperactive-distractible behaviors can be distinguished from aggressive-hostile but they are correlated because the opposite pole of each is essentially "no deviant behavior". Several scales which combine aggressive and hyperactive behavior are Miller (1972) and Bell, Waldrop and Weller (1972). Those which separate hyperactivity from aggression include Behar (1974), Schaefer and Aaronson (1966), Miller's LBCL (1973), and Kohn and Rosman (1972a) whose scale contains no items related to the hyperactive-distractible domain.

When aggression and hyperactivity-distractibility are not separated, there may be confusion regarding whether treatment influences aggressive behavior, hyperactive-distractible behavior or both. Past research with older children, for example, suggests that stimulant medication is most likely to be of value when the behavior problem is characterized by distractibility, short attention span and hyperactivity. There is considerable information concerning the poor long-term prognosis for young children with aggressive behavior problems (Conger and Miller, 1966; Glick, 1972, Robins, 1966). Glock (1972) has suggested that this poor prognosis is not associated with the syndrome that is characterized primarily by restlessness, overtalkativeness and attention-getting behavior. It may be particularly important to keep these dimensions separate in the preschool period since Aggressive-Defiant behavior and Hyperactive-Distractible (Low task orientedness) have different predictive relations with later school achievement. Kohn and Rosman (1974), for example, found high correlations between preschool ratings of low task orientedness and later school achievement but no correlations between preschool aggressive behavior and later school achievement.

The second question of whether behavior dimensions are unipolar or bipolar seems primarily related to item selection and the type of population being characterized. Bipolar scales tend to emerge from studies on normal populations where items are selected to reflect the range of most common observable behaviors. These scales may not be useful in identifying significant but rare deviant behavior (e.g., fire setting, smearing feces, suicide attempt). Such bipolar scales are reported by Schaefer and Aaronson (1966), the Kohn and Rosman Social Competence Scale (1972a) and by the Social Competence Scale of Levine, Elzey and Lewis (1969). Bell, Waldrop and Weller (1972) report an ostensibly bipolar scale with withdrawal at one end and hyperactivity-aggression at the other. However, this seems partially a result of their item selection and scoring method, as well as their choice of statistical analysis.

In monitoring behavioral change, it is desirable that at least one scale be selected that includes ratings of prosocial behavior as well as deviant behavior. Since positive and negative behaviors are not always mutually exclusive, improvement that is reflected primarily in an increase in positive behavior may be missed if only scales dealing with deviant behavior are included.

The three scales which have been selected for detailed review include Behar's PBQ (Preschool Behavior Questionnaire), (Behar, 1973), Kohn and Rosman's Social Competence Scale and Symptom Checklist (1972a), and Schaefer and Aaronson's Preschool and Primary Behavior Scale (1966). Additional scales which were examined include Chamberlin's (1974), Eisenberg, Landowne, Wilner and Imber (1962), Fels Child Behavior Scales (1941), the Bell et al. Scale (1972), and the Social Competence Scale of Levine et al. (1969). These latter scales were not considered further either because they appeared to have been supplanted by equally good or better scales, standardization was incomplete or limited, or because they were cumbersome to use.

The three scales presented all have some data on test-retest reliability and inter-observer reliability. All three scales present information concerning content validity of the scales. Criterion validity was not examined directly in the Schaefer and Aaronson work, but extensive

consideration has been given to the question of construct validity (Schaefer, 1971) and predictive validity has been studied by Kohn and Rosman (1974). Both the Preschool Behavior Questionnaire of Behar (1974) and the Symptom Checklist of Kohn and Rosman (1972 a) were developed with considerations for criterion validity, which, in the case of the Kohn and Rosman scales, includes studies of predictive validity. In view of the great similarity among these three scales in the types of behavior dimensions identified and their intercorrelations with each other, it seems safe to assume that construct and criterion validity established for one may be extended to the other.

Parent ratings of child behavior have been studied extensively, primarily by Schaefer and his colleagues and by Miller and his (1973). In Schaefer and Aaronson's (1966) scale, The Home Behavior Inventory, the preschool child's behavior is divided into six bipolar groups: Extroversion-Introversion, Hostility-Considerateness, and Task-Oriented Behaviors-Distractibility. The parent rates the child on five behaviors within each group on a five-point scale ranging from "almost always" to "almost never". The checklist takes between 5 to 10 minutes to complete. The Louisville Behavior Checklist (LBCL) (Miller, 1973) is designed to help the parent pinpoint behaviors of the child characteristic of a whole range of childhood behavior disorders. Form E1 is the 1973 revision of the Checklist appropriate for use with children age 3 to 6 years. The parent marks the items of child behavior either "T" (true) or "F" (false). A mental health worker or other professional knowledgeable in child psychopathology may need to be present to help the parent make the unqualified "T - F" judgement. Prosocial and highly pathological items have been placed at visible points to permit rapid scanning of the inventory. There are 19 scales, the first 11 of which are factor scales. The checklist can usually be completed in 1/2 hour.

Those who have written about the use of parent rating scales suggest that great caution must be applied in interpreting results particularly since parent ratings often fail to reflect changes observed elsewhere (Novick, et al. 1966, Miller, Hempe, Barrett, and Noble, 1972; Miller, 1973; Schaefer, 1971).

However, the LBCL, because it deals with a wide range of problem behaviors and includes rare behaviors, might be particularly useful in describing a population under study or searching for side effects.

Where longitudinal follow-up between preschool and school-age children is anticipated, scales which span the entire preschool-school age range will usually be preferrable. The best group of scales for this purpose are those of Schaefer and Aaronson.

Miller's LBCL for parents has a similar range as does the School Behavior Checklist (SBCL) (Miller, 1972). In the case of the SBCL, however, normative data are not yet available for the 3 to 6 year age; items are answered as "true" or "false" rather than on a scale of severity, and hyperactive-distractible behaviors are not distinguished from aggressive.

Title: PRESCHOOL BEHAVIOR QUESTIONNAIRE

Author: Lenore Behar

Age Range: Preschool

Available From: Learning Institute of North Carolina 1006 Lamond Avenue, Durham, North Carolina 27701

Located From: Behar, L., and Stringfield, S. A behavior rating scale for the Preschool child. Developmental Psychology, 1974, 10, 601-610.

Description

This scale was developed as a screening instrument to be used by teachers and child care workers in the early detection of children's emotional problems. The Preschool Behavior

Questionnaire is a modification of the Children's Behavior Questionnaire (CBQ), a 26-item behavior checklist previously standardized in England on elementary school children (Rutter, 1967). The Preschool Behavior Questionnaire (PBQ) consists of 29 items, rated on a three-point scale from "Doesn't apply" to "Certainly applies". Examples of the items are "squirmy, fidgety child", "tells lies", bites nails or fingers", and "tends to be fearful or afraid of new things or new situations". Scores are derived on three factors, Anxious, Hostile and Hyperactive.

Standardization

The normal sample of 496 children was chosen from five preschools in Durham, North Carolina, and two in Portland, Oregon. Schools were selected from various areas of the two cities so that the children represented socio-economic groups ranging from lower to upper middle class. The samples are roughly comparable to the general population in terms of numbers of white and black as well as male and female children. The emotionally disturbed sample was drawn from 15 preschools, throughout the country, that are involved in early intervention work with behavior-disturbed children. There was a sample of 102 preschoolers in this group, whose primary diagnosis was emotional disturbance. The original items in the long form were used for standardization.

Reliability

Average inter-rater reliability coefficients ranged between .67 for Hyperactive factor to .84 for Total scale. Average Test-retest reliability over a 3 to 4 month period ranged from .60 for the Anxious factor to .94 for the Hyperactive factor.

Validity

In the original study, 31 of 35 items differentiated between normal and deviant children at the .01 level. The best discriminating items identified by multiple regression analysis were similar to items which others had previously found to discriminate between deviant and normal children. In a second study on 89 children the scale was found to differentiate between normal and deviant children.

Comment

The author's purpose was to develop a scale that would be applicable to the preschool child, have standardization information on both a normal and disturbed population and be brief enough to be used as a screening tool by a teacher. Her aims seem to have been accomplished. The scale is well designed and although predictive value to later functioning has not been published, it is close enough to longer scales to make it likely that such information will be similar to that obtained with the other scales. A recent review (Behar L, 1977) presents material concerning ongoing research with the instrument.

Title: SOCIAL COMPETENCE SCALE AND SYMPTOM CHECKLIST

Authors: Martin Kohn and Bernice Rosman

Age Range: Preschool

Available From: Martin Kohn, The William Alanson White Institute
20 West 74th Street, New York, New York 10023

Located From: Kohn, M, and Rosman, B. A social competence scale and symptom checklist for the preschool child. Developmental Psychology, 1972, 6, 430-444.

Description

Both instruments focus on overt classroom behavior. The Social Competence Scale was designed to measure interpersonal functioning in the classroom. The scale consists of 90 items rated on a 7- point frequency scale (from always to never). The Symptom Checklist consists of 58 items and attempts to cover the major clinical symptoms which preschool children manifest in preschool and day care settings. The items consist of statements indicative of clinical disturbance in this age group. Two factors were identified in their pool of items. Factor I is termed Interest- participation versus Apathy-withdrawal. Factor II is termed Cooperation-compliance versus Anger-defiance. These two factors emerged from both the Symptom Checklist and the Social Competence Scale. The corresponding Factors on the two scales were highly correlated (-.75 to -.79).

Standardization

All children (n = 407) in six day care centers in the Division of Day Care in the New York City Department of Social Services were rated on the 58-item Symptom Checklist and the 90-item Social Competence Scale. Three of the centers had a primarily white population and three a primarily black population.

Reliability

Inter-rater reliabilities for pooled scores on the two scales ranged between .73 and .90 in two studies (Kohn and Rosman, 1972b). For Factor I, test-retest reliabilities over a 6 month interval ranged between .60 - .66 at 12 months, .41 - .44, and 18 months, .35 - .38. For Factor II, test-retest reliability coefficients over a 6 month interval were .73 - .77 (for the same teacher) and .54 - .59 between different teachers.

Validity

Preschool ratings of Interest-participation were related to preschool cognitive behavior (Kohn and Rosman, 1973) but ratings on Anger-defiance were not. Similarly, significant partial correlations were obtained between preschool ratings of Interest-participation and both later achievement (1974) and social-emotional functioning (Kohn and Rosman, 1972).

Comment

When combined with the Task-oriented factor from the Schaefer and Aaronson scales, these two scales have much to recommend them though it is debatable whether they offer more than the Schaefer-Aaronson Scales alone.

Title: CLASSROOM BEHAVIOR INVENTORY (FORM FOR PRESCHOOL AND PRIMARY)

Authors: Earl S. Schaefer, May R. Aaronson and Victor H. Small

Age Range: Preschool through Elementary

Available From: M. R. Aaronson, Center for Studies of Child and Family Health, National Institute of Mental Health, 5600 Fishers Lane, Rockville, Maryland 20852. For the Head Start Planned Variation version, request ERIC PN #002801 from ERIC Document Reproduction Service, Leasco Information Products, 4827 Rugby Avenue, Bethesda, Maryland 20014.

Located From: Socioemotional Measures for Preschool and Kindergarten Children, Walker, D. K., Jossey-Bass, San Francisco, 1973.

Description

The older version of this teacher rating scale measures three bipolar behavior traits: extroversion versus introversion, positive social behavior versus social hostility, and positive task-oriented behavior versus negative task-oriented behavior. The teacher rates the child on 60 items based on a 4-point scale ranging from +1.0 for "not at all like" to +4.0 for "very much like" (Schaefer, 1971). Examples of the items are: "moves from one area of the room to another frequently", "plays alone unless he's induced to play with other", "joins a group of his own accord during games, free time, etc.", and "likes to talk about everything that happens to him". For scoring purposes the 60 items are divided into 12 traits. Examples of these traits are: "verbal expressiveness," and "hyperactivity," and "concentration." A child is given a score for each trait, computed by summing the points for the trait's items. A variation of this scale rates the child on 15 seven-point rating scales, ranging from +1.0 for "never" to +7.0 for "always." The 7-point rating scale, also called the Schaefer Behavior Inventory, was used in the last year of the Head Start Planned Variation Study (Walker and others, 1973) and in the pilot year of the Home Start Study (Hi-Scope, 1973). This version measures three behavior traits: task-orientation, extroversion and hostility.

Examples of the five task-orientation items are: "stays with a job until he finishes it" and "becomes absorbed in what he is doing". Examples of the five extroversion items are: "tries to be with another person or group of people" and likes to take part in activities with others". Examples of the five hostility items are: "slow to forgive when offended" and "stays angry for a long time after a quarrel". A child is given a score for each trait or subtest, computed by summing the points for the subtest's five items. A low score represents an infrequent manifestation of the trait measured.

Standardization

Means and standard deviations for each of the three subtests for the total fall 1971 Head Start Planned Variation sample (n = 4943) and subsamples (males, females, black children, white children, Mexican-American children, children with previous preschool experience, and children with no previous preschool experience) are available (Walker and others, 1973).

Reliability

Factor analytic studies, using a principal component analysis, with two preschool samples revealed three distinct independent factors representing each of the three traits. The three factors that emerged from an analysis of the 464 Head Start children's scores in a reliability study from the Head Start Planned Variation Study explained 80.4% of the total variance in the Home Start pilot study with 173 children (Hi-Scope, 1973). Ceiling and floor effects in the distribution of the three subtest scores were found in both preschool studies. Test- retest reliability coefficients after a 3-week interval were in the .70's for a sample of 464 Head Start children in 4 sites (Walker and others, 1973). Inter-rater reliability coefficients (product-moment correlations and Spearman rank-order correlations) between classroom aides and other paraprofessionals in 13 Head Start Planned Variation classrooms were highest for the task orientation scores (medians .62 and .60) and lowest for the extroversion scores (medians .46 and .49) and for the hostility scores (medians .39 and .44). There were large discrepancies between mean scores for the two raters, especially in the extroversion Scores (Walker and others, 1973). No other inter-rater reliability estimates are available for analysis purposes. Inter-rater reliability (coefficient alphas) estimates calculated for the Home Start pilot sample were .72 for task orientation, .72 for extroversion and .67 for hostility. Item analysis in the same study revealed that an item correlated higher with its two scales (Hi-Scope, 1973).

Validity

The four-point scale version of the inventory was used to assess adjustment in a study with 134 Mexican-American Head Start 5- year-old enrollees in Texas. The correlation of the

standardized inventory total scores with the Tests of Basic Language Competency was .40 for the English version and .08 for the Spanish version (Stedman and McKenzie, 1971).

Kohn and Rosman (1972) found high correlations between their own scales and the Schaefer-Aaronson Introversion-Extroversion factor and the adjustment-maladjustment factor. For follow-up they pooled scores from their own scales with scores from these two Schaefer and Aaronson factors. The predictive validity of Schaefer-Aaronson Task-Oriented factor was studied alone. Preschool ratings on the pooled scores representing Interest- Participation-Extroversion and on Task-orientedness alone showed significant partial correlations with all measures of achievement at the end of second grade. Anger-defiance in the preschool period was not related to later school achievement.

Comment

This is probably the single most versatile and complete scale available. In addition to the preschool form there is an upward extension to the school aged child. Because of similarity in items with other scales, it is probable that deviant scores on this scale will identify deviant children. However, identification of deviance has not been studied directly and the scales do not cover rare behaviors. Supplementation of this scale with either the Behar Scale or the Kohn and Rosman Symptom Checklist would be advisable when used to define degree of pathology in a child.

REFERENCES

Behar, L., & Springfield, S. A behavior rating scale for the preschool child. Developmental Psychology, 1974, 10, 601-610.

Behar, L. The preschool behavior questionnaire. J. of Abnormal Child Psychology, 1977, 5, 265-276.

Bell, R., Waldrop, M., & Weller, G. A rating system for the assessment of hyperactive and withdrawn children in preschool samples. American Journal of Orthopsychiatry, 1972, 42, 23-34.

Buros, O. K. (Ed.) The seventh mental measurements yearbook. (Volume I). New Jersey: The Gryphon Press, 1972.

Chamberlin, R. The use of teacher checklists to identify children at risk for later behavioral and emotional problems. Rochester, New York: University of Rochester Medical Center, 1974.

Conger, J. J., & Miller, W. C. Personality, social class and delinquency. New York: John Wiley & Sons Inc., 1966.

Eisenberg, L., Landowne, E. J., Wilner, D. M., Imber, S. D. Children's guild symptom checklist and children's guild health inventory. American Journal of Public Health, 1962, 52, 18-28.

Fels Child Behavior Scales. Yellow Springs, Ohio: Fels Research Institute, 1941.

Glick, S. J. First follow-up study of Glueck table to identify predelinquents at school entrance. In S. Glueck & E. Glueck (eds.) Identification of predelinquents. New York: Intercontinental Medical Book Corporation, 1972.

Johnson, O. G., & Bommarito, J. W. Tests and measurements in child development: A handbook. San Francisco: Jossey Bass, 1971.

Kohn, M., & Rosman, B. Relationship of preschool social- emotional functioning to later intellectual achievement. Developmental Psychology, 1972, 6, 445-452.

Kohn, M., & Rosman, B. A social competence scale and symptom checklist for the preschool child: Factor dimensions, their cross-instrument generality, and longitudinal persistence. Developmental Psychology, 1972, 6, 430-444.

Kohn, M., & Rosman, B. Cognitive functioning in five-year-old boys as related to social-emotional and background-demographic variables. Developmental Psychology, 1973, 8, 277-294.

Kohn, M., & Rosman, B. Social-emotional, cognitive, and demographic determinants of poor school achievement: Implications for a strategy of intervention. Journal of Educational Psychology, 1974, 66, 267-276.

Levine, S., Elzey, F., Lewis, M. California Preschool Competence Scale. Palo Alto, California: Consulting Psychologists Press, Inc., 1969.

Miller, L. C. School behavior checklist: An inventory of deviant behavior for elementary school children. Journal of Consulting and Clinical Psychology, 1972, 38, 134-144.

Miller, L. C. Louisville Behavior Check List Form E1. Mimeographed manual available from Lovick C. Miller, Child Psychiatry Research Center, University of Louisville School of Medicine, Louisville, Kentucky.

Miller, L. C. Dimensions of psychopathology in middle childhood. Psychological Reports, 1967, 21, 897-903.

Miller, L. C., Hempe, E., Barrett, C. L., & Noble, H. Children's deviant behavior within the general population. Journal of Consulting and Clinical Psychology, 1971, 37, 16-22.

Miller, L. C., Hempe, E., Barrett, C. L., & Noble, H. Test-retest reliability of parent ratings of children's deviant behavior. Psychological Reports, 1972, 31, 249-250.

Novick, J., Rosenfeld, E., Block, D. A., & Dawson, D. Ascertaining deviant behaviors in children. Journal of Consulting Psychology, 1966, 30, 230-238.

Robins, L. N. Deviant children grown up. Baltimore, Maryland: Williams & Wilkins, 1966.

Schaefer, E. S. Development of hierarchical, configurational models for parental and child behavior. J. P. Hill (ed.), Minnesota Symposia on Child Psychology, Volume V. Minneapolis, Minnesota: University of Minnesota Press, 1971.

Schaefer, E. S., & Aaronson, M. R. Classroom behavior inventory: Preschool to primary. Bethesda, Maryland: National Institute of Mental Health, 1966.

Schleifer, M., Weiss, G., Cohen, N., Elman, M., Cvejic, H., Kruger, E. Hyperactivity in preschoolers and the effect of methylphenidate. American Journal of Orthopsychiatry, 1972, 6, 430-444.

Walker, D. K. Socioemotional measures for preschool and kindergarten school. San Francisco: Jossey-Bass, 1973.

APPENDIX VIII

SELF REPORT MEASURES*

As self report measures have been of key importance in psycho-pharmacologic studies with adults, it is reasonable to consider whether similar ratings are possible and useful with pediatric populations. It will be evident that the concept of a "self report" is at some point an arbitrary distinction from that of a structured interview or of some projective tests, particularly with younger populations who do not read. Any of the "self report" measures discussed here might be better administered as a questionnaire-interview with the testor recording the responses. The technique of administration should be carefully specified. The appendices on Observational Techniques and on Social and Emotional Assessment should also be consulted for some possible overlap of material.

The area of self report measures is less well developed than others described in these guidelines, and the literature on self report with children is limited to a handful of "personality" measures or "self concept" scales which have only preliminary data concerning reliability and validity, and some limited work on the self rating of "anxiety" or "depression" in childhood. These are briefly surveyed here.

I. Personality Ratings

Personality "inventories" were the earliest form of self report scales used with children. They aim at general reports about feelings and behavior in a variety of situations based on a general notion about "adjustment" or "happiness."

The Personality Adjustment Inventory (Rogers, 1931) is recommended for possible research potential. This test was first published in 1931 under the title of "Test of Personality Adjustment," and was reissued but not revised in 1961, under the new title. There are six parts to this test, each using a different approach in assessing a child's attitudes toward self, family, and peers (see copy of test in Appendix). While designed for ages 9-13, the wording and sensitivity of some of the items might be appropriate for younger populations. The areas covered are labled "personal inferiority, social maladjustment, family maladjustment, and daydreaming." Scoring is cumbersome, however, and the test is not presently recommended as a psychometric instrument. The test has not been used in any psychopharmacologic study to date.

In spite of the unsatisfactory validity and reliability data (see Smith, 1958), this test may be the best approach to an age group for whom relatively indirect methods will be the most satisfactory in maintaining the child's interest and cooperation. Both the language and concepts need to be updated, however, and there is considerable work to be done with this measure.

The Children's Personality Questionnaire (CPQ) was devised by Raymond Cattell and colleagues for use from 6 years of age to adulthood, to measure several behavioral characteristics (Porter and Cattell, 1959). Fourteen scores are obtained for entities labeled: reserved vs. warmhearted, sober vs. happy-go-lucky, relaxed vs. tense, etc. Normative data and reliability data are given, yet it is not clear how to relate the 14 independent factors which Cattell extracts from these scores to clinically meaningful

*By J.L. Rapoport, M.D.

concepts, and clinical validating data is lacking. As recommended in the 1963 revision of the Questionnaire, a total of 280 items should be completed by the child, making this a relatively lengthy examination. The test has not been used in any pharmacologic studies, but might have application as a research instrument.

The MMPI has not been used extensively with preadolescent populations; however, selected scales have been constructed from it. A 26-item Social Desirability Questionnaire has been reported for use with nursery school children. Test retest reliabilities were moderate for children over four, and validity studies suggest that children scoring higher on the scale were more motivated to respond positively to interpersonal demands (Ford and Rubin, 1970). The Psychopathic-Deviant Scale of the MMPI has been found related to refractory school behavior of young adolescents (Davies & Maliphant, 1971) and might be useful with preadolescent populations.

II. "Self Concept" Scales

Some considerable research has gone into the notion of "self concept," particularly in relation to "ideal self" measures. Piers and Harris (1964) developed a 30-item self concept scale derived from factor analysis of an initial 140-item scale. Children are asked to answer "yes" or "no" to statements about themselves, such as "I do many bad things," or "I have a pleasant fact." This scale was assembled in a rational manner and low but significant correlations were demonstrated between ratings of "self concept" and IQ and academic achievement. Clinic children, age 8-14, had slightly but significantly lower scores on the scale than did age matched controls (Piers, 1972). In this author's experience, this scale proved uninviting for a population of hyperactive boys, but some (low) correlations were obtained between changes on this measure and changes in school behavior and performance over a six-week period (Rapoport et al., 1974). The scale is discussed here because of the paucity of other rating instruments in this area, and because it might prove more appetizing to internalizing populations than it did to behavior disordered groups. Nichols and Berg (1970) constructed a 15-item modification of the Piers and Harris scale, which utilized a semantic differential task. Reliable self concept ratings from a group of school phobic children were reported. There is other evidence to suggest that the semantic differential is suitable and reliable for use with children (DiVesta and Dick, 1966).

A variety of other self concept scales have been employed (Havighurst, 1946; Bills et al., 1951), which have shown some relation to scholastic ability (Bailey et al., 1974). A similar approach is reported by Bower (1969) for a Self Rating Scale (grades 3-7) which has the child first rate 40 attributes concerning a "desired" self and then in a second part rate 40 items of a "perceived" self. This instrument is reported as a preliminary research attempt, is not yet a psychometric instrument, but may have research promise, particularly with "depressed" populations.

Simmons and Rosenberg (1971, 1973) have presented a scale for a total "self inventory" which was utilized by them for sociological evaluations of large urban school populations (grades 3-12). The total self-inventory consists of six sections: self consciousness, stability of self, self-esteem, perceived self image, affect and self image. Data on test construction is not given, and validity measures, although significant, are low. Scoring is ambiguous and difficult to interpret; in a recent study, all scales failed to distinguish a group of hyperactive boys from controls even when academic failure and independent measures of peer rejection did successfully discriminate these groups (Riddle and Rapoport, 1975).

Quay and Peterson (1958) reported a 40-item true-false self rating scale which was reliable for a 7th-8th grade population of delinquents and which distinguished 67% of delinquents from controls (total N from three studies = 781). The scale might be useful for younger populations.

III. Anxiety and Mood Ratings

The children's Manifest Anxiety Scale, CMAS (Castenada, et al., 1956) for use with 4th through 6th grade populations, consists of 42 items taken from the Taylor Manifest Anxiety Scale (Taylor, 1953), such as "I am nervous," or "I worry when I go to bed at night," and an 11-item lie scale with items such as "I like everyone I know". Scores are the total count of "yes" responses unpopularity and lower academic achievement (McCandless, 1967). Sarason, et al. (1958) have developed a 30-item "yes-no" questionnaire for children (grade I and up) about test anxiety. Children must respond to items such as, "Do you worry a lot before a test?" The scale has demonstrated adequate test retest reliability and low positive correlations with teacher rating of anxiety (Sarnoff, et al., 1959). Both this test and the CMAS have received somewhat favorable reviews as potential research instruments (Bronfenbrenner and Ricciuti, 1960). A self rating of anxiety may be clinically interesting with children presenting with phobic or anxiety symptoms; other clinical populations may provide more "anxious" profiles on the basis of general immaturity, and therefore high scores may not necessarily be specific for anxiety states.

No self rating of "mood" per se is available for preadolescent populations, although affective state is often assessed by direct observation. One preliminary attempt has recently been reported, however (Kovacs and Beck, 1975), in which a wide range of positive responses to depressive items was demonstrated for 7th and 8th grade children. It may be that such a rating instrument will become available in the near future.

IV. Global Self Reports (highly recommended)

It should be noted that global self ratings by children have rarely been reported. In one study (Gittelman- Klein and Klein, 1971), children's global ratings of improvement distinguished drug from placebo (100% compared with 24%) better than did any other individual measure, and such self reports should be obtained routinely.

V. Issues and Problems

There are several important issues concerning self report measures in children, particularly for younger groups. The method of administration of the scales may need to be adapted for a particular group. Several workers find that children below 4th grade cannot be relied upon to complete paper and pencil self-administered "personality inventories;" some of the self report techniques described here, however, could be used with younger children if read aloud and recorded by an interviewer. Even with careful administration, it is not always clear whether the unsatisfactory data self report scales may elicit occur because of the child's inadequate conceptualization of his or her difficulty, or because the rating scales are actually addressing themselves to entities which are not valid constructs for that particular age group. This is a particularly controversial question when one considers self-ratings of depression, as it is a matter of considerable debate whether or not depression, in the adult sense, actually occurs in childhood.

A second and related issue is whether direct self reports are appropriate measures at all, particularly for children under 8 years of age. For these reasons, most investigators have been cautious about using verbal reports because of these many sources of invalidity, i.e., lack of self awareness, inability to conceptualize and verbalize, and, on the investigator's part, the presumed existence of the underlying variable.

The tests discussed here may lack both the flexibility to record what actually transpired during the self report process and also the rigor of other ratings. Nevertheless, for clinical pharmacologic studies, a strong recommendation is made to include some self report measures. The caretakers of the child (parent, teacher,

and/or ward personnel) will hear and record children's self reports, and these may have considerable impact upon clinical planning in some settings. For this reason, therefore, even negative results provide some guidelines where none presently exist for the treating clinician. Secondly, pediatric psychopharmacologic studies to date have focused on behavior disordered, schizophrenic, or retarded populations. If future studies deal more with anxiety or depression in childhood, then it is probable that self report measures will have considerably more relevance and sensitivity than they have had for the former populations.

VI. Suggestions for Future Research

Individualized "self report" scales need to be developed for specific studies, i.e. target symptom self reports in which a child with a given problem is asked to rate his specific problem in its setting (such as fear of separation, fear of going to bed) during the course of a study on a scale agreed upon at the outset. Such ratings would allow greater flexibility while still permitting critical input from the child. Both global assessments and specific items of change should be elicited from the child to distinguish effect on mood from perception of specific changes.

"Lie" scales such as that on the CMAS have obvious shortcomings (Rie., 1963) but may prove useful as measuring instruments in themselves. As middle childhood is an age in which a certain amount of denial is expected and healthy, the Lie scale capitalizes on an abundantly occurring interview behavior. In an unpublished study of behavior disordered grade school boys, this author found the Lie scale to be significantly correlated with school indices of conduct disorder for 8 and 9 year old boys, while other clinic measures did not.

For younger populations (age 7 and under), displaced ratings (completion of story, questions about wishes and doll play) may prove more fruitful than self report, and the reader should see the review by Yarrow (1960) for alternate techniques, as well as the discussion of the Playroom Interview in the Appendix on Social and Emotional Assessment.

REFERENCES

Bailey, R.C., and Bailey, K.G.: Self-perceptions of scholastic ability at four grade levels. J. of Genet. Dev., 124, 197-212, 1974.

Bills, R.E., Vance E.L., and McLean, O.S.: An index of adjustment and values. J. Consult. Psychol., 15, 257-261, 1951.

Bower, E.M.: Early Identification of Emotionally Handicapped Children in School. 2nd ed., Springfield, Ill., Charles C. Thomas, 1969.

Bronfenbrenner, U. and Ricciuti, H.N.: The appraisal of personality characteristics. In P.H. Mussen (Ed.): Handbook of Research Methods in Child Development. John Wiley and Sons, New York, 1960.

Castenada, A., McCandless, B.R.: The children's form of the Manifest Anxiety Scale. Child Dev., 27, 317-325, 1956.

Davies, J.G. and Maliphant, R.: Refractory behavior at school in normal adolescent males in relation to psychopathy and early experience. J. Child Psychol. and Psychiat., 12, 35-42, 1971.

DiVesta, F.J. and Dick, W.: Test-retest reliability of children's ratings on the semantic differential. Educ. Psychol. Measur., 26, 605-616, 1966.

Ford, L.H. and Rubin, B.M.: A social desirability questionnaire for young children. J. Consulting and Clin. Psychol., 12,35-42, 1971.

Gittelman-Klein, R., and Klein, D.F.: Controlled imipramine treatment of school phobia. Arch. Gen. Psychiat., 25, 204-207, 1971.

Havighurst, R.H.: The development of the ideal self in childhood and adolescence. J. Educ. Res., 40, 241-257, 1946.

Koocher, C.P.: Emergine selfhood and cognitive development. J. Genetic Psychol., 125, 79-88, 1974.

Kovacs, M. and Beck, A.: An empirical-clinical approach towards a definition of childhood depression. Paper presented at the NIMH Conference on Childhood Depression, Sept. 19-20, 1975, Washington, D.C.

McCandless, B.R.: Children, Behavior and Development 2nd ed., Hinsdale, Ill., Dryden Press, 1967. Chap. 6, 253-292.

Nichols, K. and Berg, I.: School phobia and self-evaluation. J. Child Psychol. and Psychiat., 11. 133-141, 1970.

Piers, E.V.: Parent prediction of children's self concepts. J. Consulting and Clinical Psychology, 38, 428-433, 1972.

Piers, E.V., and Harris, D.B.: Age and other correlates of self concept in children. J. Educ. Psychol., 55, 91-95, 1964.

Porter, R.B., and Cattell, R.B.: Children's Personality Questionnaire. Illinois, Institute of Personality and Ability Testing, 1959.

Quay, H. and Peterson, D.: A brief scale for juvenile delinquency. J. Clin. Psychol., 14, 139-142, 1958.

Rapoport, J.L., Quinn, P.O., Bradbard, G,, Riddle, K.D., and Brooks, E.: Imipramine and methylphenidate treatments of hyperactive boys. Arch. Gen. Psychiat., 30, 789-791, 1974.

Riddle, K.D., and Rapoport, J.L.: A two year follow-up of 72 hyperactive boys. J. Nerv. Ment. Dis., Vol. 162, 1976, 126-134.

Rie, H.: An exploratory study of the CMAS lie scale. Child Development, 34, 1003-1017, 1963.

Rogers, C.R.: Measuring Personality Adjustment in Children, Six to Thirteen Years of Age. Columbia University Press Contributions to Education No. 458; New York, Publications of the College, AMS Press, 1931.

Rosenberg, M. and Simmons, R.: Black and White Self Esteem: The Urban School Child. The American Sociological Association, Washington, D.C. 1971.

Sarason, S.R., Davidson, K.S., Lighthall, F.K., and Waite, R.R.: A test anxiety scale for children. Child Dev., 29, 105-114, 1958.

Sarnoff, I., Sarason, S.B., Lighthall, F.K., and Davidson, K.S.: Test anxiety and the "eleven plus" examination. Brit. J. Educa. Psychol., 49, 129-136, 1959.

Simmons, R.G., Rosenberg, F.: Disturbance in the self image at adolescence. Amer. Sociolog. Review, 38, 553-568, 1973.

Smith, L.: The concurrent validity of six personality and adjustment tests for children. Psychol. Monograph 72, 1-30, 1958.

Taylor, J.A. A personality scale of manifest anxiety. J. Abnorm. Soc. Psychol., 48, 285-290, 1953.

Yarrow, M.: The measurement of children's attitudes and values. In Mussen, P. (Ed.): Handbook of Research Methods in Child Development. New York, J. Wiley, 1966, Chap. 16, 645-687.

APPENDIX IX

SOCIAL AND EMOTIONAL FUNCTIONING*

INTRODUCTION

It is important to distinguish between the diagnostic - evaluative process necessary for decision for clinical disposition to be made, and the attempt to systematize individual measures of the child's social and emotional functioning. This discussion will focus on individual measurement for portions of the diagnostic clinical assessment which remains a complex judgment concerning the nature and severity of presenting complaints and symptoms, their duration and degree of disturbance in the functioning of the child, family and/or community. In general, such an assessment must involve a judgment about that individual's functioning relative to what is appropriate for his age, utilizing parallel evaluations of the nature and quality of caretakers who are reporting about the child. Reports must also be assessed within the broader framework of the school, community and geographic area in which the assessment is made.

While this survey contains some recommendations for systematized measurement in individual areas of data collection, so far the best approach to the child as a patient is through careful evaluation of those unique individual aspects of any given case. A general reference to familiarize the researcher with the methodologic problems inherent in assessment of the child's interpersonal functioning is found in Rutter et al. (1970), whose landmark epidemiologic study provided the first comprehensive overview of the strengths and weaknesses of the diagnostic process in child psychiatry. This appendix will cover measurement of social and emotional functioning of the child through reports from parents, teachers, peers, psychological testing and playroom interview. It is hoped that further development of such measuring instruments will close the gap between the clinical process and research-based measures. Other appendices (self-reports, assessment of family and environment) will necessarily overlap this report and should also be consulted.

I. Information From Parents

Rutter et al. (1970) have shown that information from the parents is the most critical single factor in the diagnostic process, in spite of the difficulties of formalizing such information. No one else has the amount of material concerning the child in so many different situations and over such a long time span. However, the inaccuracy of retrospective behavioral data has been well documented (Yarrow et al., 1970), and there is no systematized interview or form for recording historical data in general clinical use.

While few attempts to formalize diagnostic interviewing with parents have been made, most clinicians agree that it is useful to see both parents together for part of the interview, and the parents together with the child, at least for brief periods. To record basic data in a uniform manner about the family background, developmental history of the child, and some individual symptoms, two rating instruments were developed for the ECDEU assessment battery (Psychopharmacology Branch, NIMH): the Children's Personal Data Inventory (CPDI) and the Children's Symptom History (CSH). However, these instruments will not be sufficient for any given study; the content of these forms

*Written by J. Rapoport, M.D.

is biased towards outpatient studies with nonpsychotic children, and so provide both too narrow a range as well as insufficient information with regard to any specific target symptom. Behavioral history will need amplification, for example, with quantification of target symptoms at baseline and during treatment. Clinically, a judgment is always made concerning the parents' reason for seeking consultation, their expectations from the evaluation or study, and their own theories about the cause of the problem. There is presently no rating scale for recording this information, but such a scale could easily be devised, and would be useful in interpreting other information from the parents.

Background information from the parents should also include data about the child's way of functioning in a variety of specific situations. Longitudinal studies (Thomas et al., 1968; Rutter, 1964; Graham, Rutter and George, 1973) have indicated that specific styles of behavior or "temperament" (regularity, rythmicity, mood, threshold, activity, persistence, approach, adaptability, intensity) have some predictive value in identifying children "at risk" for behavior disorders. While the Thomas and Chess study evaluated "temperament" from lengthy interview, they were able to validate this material by direct observation with a subgroup of their sample.

A rating scale has recently been reported to indicate "temperament" for middle childhood (Garside et al., 1975), and global evaluations of these behaviors can also be made from interview; a "temperament" scale has recently been reported for use with infants (Carey, 1970).

The parent interview is critical for determining the degree to which given target behaviors interfere with child or family functioning. One technique for evaluating specific target behavior is a daily diary in which hourly notes over a 2 - 4 day period are kept. Diary-reported behavior has been shown to correlate satisfactorily with direct observation (Rapoport and Benoit, 1975) and can be drug sensitive (Rapoport et al., 1974). Furthermore, diaries provide clinical richness, yet can be scored relatively easily, if the parents are carefully instructed as to which behaviors to record.

Clinically, judgments about family interactions are of theoretical interest, and may be critical in diagnosis and planning for the child. It is a commonplace that some behavior problems result from parental rejection or conflict, or may disappear when parents are helped to set firm limits. There is, however, no systematized observational technique in general clinical use for making such judgments.

One possible measure of parent functioning is the Revealed Difference Test (Farina, 1960), in which parents are asked to discuss the handling of hypothetical behavior problems separately, and then to resolve their differences in approach. The interpretation of this interaction is open to questions (Hetherington, 1972) but the test seems a promising research instrument for family study.

Bell (1964) has outlined the need for structured observations of parent-child interaction in order to obtain relevant data. While direct observations of family interaction are time consuming and expensive, in some research settings they might be feasible. Patterson et al. (1969) have devised a coding system for observing the child's interaction with his family, in which 33 discrete behavioral characteristics measure both target behavior of the child (tantrum, play, self stimulation, etc.) and the behavior of adults and peers (ignore, command, physical negative). The coded behavior provides not only what the child did, but what behaviors others exhibit preceding and following his behavior.

Schulman et al. (1962) utilized a structured playroom task for parents and child in which parents must keep their child playing with relatively undesirable toys, and then have the child complete stories together with them during a 45 minute period. Significant differences were obtained between the parental behavior of conduct problem and non-conduct problem children. Similarly, Campbell (1975) found significant differences in maternal behavior during a structured problem-solving situation in which mother was

present and permitted to "coach", when groups of normal, hyperactive, and learning-disabled boys were compared.

Pharmacologic intervention provides one way of investigating the "direction of effect" (Bell, 1968) as prospective sequential behavioral evaluations may allow interpretation as to what is cause and what is effect in parent-child interaction. It would be of considerable interest, therefore, if one of the above parent-child interactional measures were included in Phase II or III psychopharmacologic studies. For example, reported aggressive and rejecting parental behavior (found more frequently for behavior disordered than control groups) may decrease for the treatment groups more than for controls; overprojective or intrusive maternal behavior might diminish when the child receives an effective antianxiety agent.

In the absence of more precise observational measures of family functioning, global ratings should always be made of the ability of the home to accept and support the child. The basis for ratings must be clearly specified with clinical examples for each rating score.

II. Interview With Teacher

While teacher ratings are relatively reliable and valuable in assessing the child's functioning, frequent "false positives" are to be expected when utilizing any practicable rating scale cutoff point. A full clinical assessment, therefore, will always require some additional communication with the school, e.g., what efforts the school has made to deal with problem behaviors, and how consistent the difficulty has been during the term and during the school day. Information about the teacher's level of experience, and availability of other resources (reading teacher, school counselor, work-study programs, etc.) may account for a pattern of referral for clinical services, and may be critical in evaluation of children whose difficulties are confined to the school setting. Teacher's attitude toward a proposed treatment should be systematically recorded. Clinical attitude toward medication has been found to influence clinical outcome for adults; no study has yet assessed such attitudinal factors in relation to children's clinical drug response.

III. Peer Rating (strongly recommended)

No evaluation of the school age child is complete without information about peer relations. The symptom of "trouble getting along with other children" is strongly associated with general maladjustment independent of symptom pattern (Mensh et al., 1959) and difficulty in relationships with other children in grade school may predict problem behavior in early adult life (Roff, 1961).

Several rating scales are available for classroom sociometry (Bower, 1969; Tchechtelin and Amatora, 1944; Moreno, 1951). In outpatient settings, these present obvious methodologic difficulties, but may be appropriate for studies in special educational or institutational settings.

The Personality Rating Scale (Tchechtelin and Amatora, 1944), for use with grades 4 to 12, instructs students to rate each other on a 10-point scale on 22 traits such as pep, intelligence, sociability, nervousness, neatness, etc. Considerable consistency between (child) raters has been demonstrated, but no retest reliability or validity has been reported.

Several issues are raised by direct rating scales: these assessments may not be superior to indirect measures, and it is possible that adult interest in such measures would encourage nonsupportive relationships between peers. A promising indirect peer measure is "The Class Play" (Bower, 1969). In this scale students must nominate classmates for

20 hypothetical roles in a play. A second section of the Play elicits from each pupil roles he would prefer. This scale is recommended by the author for grades 7-12, but could probably be used with younger children. The "Class Play" was found to be drug sensitive (Elsenberg et al., 1963) with a young adolescent population.

Because of administrative issues with group ratings, it will often be more practical to obtain teacher estimate of peer relations; teacher judgments of social acceptance have been shown to correlate with sociometric ratings (Gronlund, 1959) and global teacher judgments on a 5-point scale were successful in discriminating hyperactive from control preadolescents in a 2 year follow-up study (Riddle and Rapoport, 1976).

IV. Psychological Testing

A. Projective testing

Projective testing was constructed on the assumption that such tests may tap the child's "inner world", revealing feelings and desires of which the child may not be aware, and that projectives can be useful in obtaining such information in relatively brief evaluations. Interest in projectives stems from the common clinical impression that much overt behavior may be defensive in nature, that is, that behavior may represent a reaction to distressing fantasies or preoccupations, so that a given behavior pattern, such as hyperactivity, may have several possible psychological determinants which projective testing can help to identify.

Several reviewers have discussed the use of projective testing with children in useful overviews (Henry, 1960; Miller, 1960; O'Leary, 1972). Specific projective techniques include associative tasks (such as the Word Association Test), construction and completion tasks such as the TAT, and the least structured, play techniques, such as the World Test (Lowenfeld, 1939).

The TAT (Morgan and Murray, 1935) contains 20 cards for children and is probably the best known and most widely utilized projective test in current use. Subjects are instructed to interpret the action, to tell what the characters are thinking and feeling, and to give imaginary constructions of preceding events and outcomes. A variety of abbreviated TAT tests for children, utilizing selected cards, are available (Mundy, 1972).

The Children's Apperception Test (CAT) (Haworth, 1966), consists of animal pictures for subjects ages 3 to 10; this test was devised on the questionable assumption that children might better identify with animals than with the persons on the TAT cards. The Rorschach has undergone some standardization as a diagnostic tool with children (Ames, 1952; 1959), and a variety of other projective tests, initially prepared for adults, have been adapted for use with children (e.g. Rosenzweig et al., 1948).

Circumscribed measures derived from projective testing have been somewhat related to overt behavior. Reports have claimed correspondence between Rorschach indices and behavior, but these relationships have not been particularly strong. The Rorschach is not recommended for routine use, but may qualify as a research instrument (Draguns et al., 1967; Haley et al., 1967). Aggressive incidents in TAT projective stories were significantly correlated with ratings of aggressive behavior by ward attendants for 9-15 year old boys (Mussen and Naylor, 1954) or with teach ratings of amount of fighting with younger populations (Kagan, 1956). In general, however, the validity of projective test measures has also been disappointing (Mursteln, 1963; Zubin et al., 1965), and even where some validity has been demonstrated, the same information may be available by simpler, more direct means. Standardized interviews utilizing some projective questions may be more flexible yet more reliable and valid than projective testing per se.

In spite of these reservations, the question of the relationship between fantasy and overt behavior remains an interesting research question in itself (McClelland, 1966). Psychopharmacologic studies might employ selected projectives as part of phase II or III studies to examine the effect of, for example, mood altering drugs on fantasy, and the relation between mood change as assessed by direct observation and changes in affective content of projective stories.

B. Evaluation of behavior during psychological testing

In addition to psychometric test scores (see Appendix on Cognitive Testing, by Sprague) the child's behavior during testing has considerable usefulness in clinical assessment. Psychological testing provides a structured, prolonged observational setting which may permit the most sensitive direct observations in outpatient studies. Richman et al. (1975) found behaviors during developmental testings distinguished "problem" and control three year olds. Restless behavior during psychometric testing has correlated significantly with classroom and home behavior for grade school children (Rapoport and Benoit, 1975) and has been found drug sensitive in double blind outpatient studies of hyperactive children (Rapoport et al., 1974; Klein et al., 1976). An 18-item rating scale of behavior during testing was developed for the Collaborative Project NIMCHS (1973 Bulletin) and considerable data for seven year olds will shortly be available. Additional test session measures might profitably include ratings of testor behavior, i.e., the frequency of encouragement, setting limits, etc., which may be elicited by the child and which might change during treatment. This would add unique interactional data, not easily available elsewhere.

V. Playroom Interview

Diagnostic interview with the child may be useful in assessing the child's relatedness and mood, and with psychotic children may be critical for diagnosis. In spite of this, there has been little written about interviewing children in a systematic way. In part, this gap exists because of the difficulty in standardizing an interview over an age range where children may vary widely in their developmental level and therefore their willingness and/or ability to deal directly with issues of interest to the interviewer. Yarrow (1960) has surveyed the methodological issues involved in interviewing children in an excellent review.

Some clear clinical reports have suggested categories of behavior to be noted in diagnostic interview (Beiser, 1962; Goodman and Sours, 1967) or provided an outline for a play interview with young children (Werkman, 1965). Rutter and Graham (1968) have described a half hour interview in detail, for use with 7 to 12 year old children, and demonstrated generally adequate reliability and validity for many behaviors such as emotional responsivity, verbal productiveness, and restlessness. However, they note the difficulty in obtaining reliable assessment of some items of behavior which could be quite transient, such as anxiety or depression, tearfulness or tension. Interrater reliabilities for these behaviors were adequate when rated during the same interview, but were less so between different interviews one month apart.

The Children's Psychiatric Rating Scale (CPRS) was developed by Drs. Werry, Fish, Klein, and Gittelman-Klein, for the ECDEU, based in part on the Rutter interview. The CPRS contains 63 items, each rated on a 6-point scale. It is evident that not all items will be appropriate for any given population, and the scale still leaves the setting, nature of interview (play, projective, direct questions), and duration of the interview unspecified. Nevertheless, it represents a first attempt to standardize such a rating instrument for general research use.

Dr. L. Cytryn (1975) has used the CPRS with a pilot sample of 25 outpatients, ages 5-15, with interrater reliabilities of .70 and above for items of interest to that particular study (anxiety, depression, depressed demeanor). One recent double-blind study of 6-12 year old hyperactive children (Gittelman-Klein et al., 1976) found 9 items (of 17 items used) from the CPRS to be drug sensitive. This latter finding is particularly impressive as the population in that study did not exhibit particularly deviant behavior in the one-to-one interview setting at baseline.

There are no standardized interview procedures for more seriously disturbed populations; a promising scale is the Children's Minimal Social Behavior Scale (CMSBS) (Ulmer, 1967, 1968). This is a rating of short lifelike social challenges given during a 5-minute period. Interrater reliability and diagnostic power were satisfactory. This scale is recommended for studies with psychotic and retarded populations.

The diagnostic interview with the child needs further definition and systematization, particularly with younger children. The setting needs to be specified; and nature of interview questions provided, such as a compilation of projective play settings, questions, or stories in clinical use (e.g., Despert, 1940) for utilization during interviews with young children will be necessary if the psychiatric rating is going to provide useful data for children under 7. Interview rating of target behaviors (e.g. depression, preoccupation with anxiety, restlessness, abnormal movements) for all age groups should be amplified as appropriate; in spite of the probable relative drug insensitivity of the psychiatric interview (Rapoport et al., 1974), a careful interview assessment may be as useful for its negative findings as for positive ones, as clinicians need to know when their common observations are <u>not</u> likely to be helpful in monitoring drug treatment.

REFERENCES

Ames, L.B.: Further check on the diagnostic validity of the Ames danger signals. J. of Projective Techniques, 23, 291-298, 1959.

Ames, L.B., Learned, J., Metraux, R.W. & Walker, R.N.: Child Rorschach Responses: Developmental Trends from Two to Ten Years. New York: Hoeber-Harper, 1952.

Beiser, H.: Psychiatric diagnostic interviews with children. J. Am. Acad. Child Psychiat., 1, 656-670, 1962.

Bell, R.Q.: Structuring parent-child interaction situations for direct observation. Child Dev., 35, 1009-1020, 1964.

Bell, R.Q.: A reinterpretation of the direction of effects in studies of socialization. Psychol. Rev., 75, 81-95, 1968.

Bower, E.M.: Early Identification of Emotionally Handicapped Children. Springfield, IL, Charles C. Thomas, 1969, 174-175.

Campbell, S.: Mother-child interaction: a comparison of hyperactive, learning disabled, and normal boys. Amer. J. Orthopsychiat., 45, 51-56, 1975.

Carey, W.B.: A simplified method for measuring infant temperament. J. Pediatrics, 77, 188-194, 1970.

Cytryn, L.: Personal communication, 1975.

Despert, J.L.: A method for the study of personality reactions in preschool age children by means of analysis of their play. J. of Psychol., 9, 17-29, 1940.

Draguns, J.G., Haley, E.M. and Phillips, L.: Studies of Rorschach content: a review of the research, Part I: Traditional contest categories. J. of Projective Techniques and Personality Assessment, 31, 3-32, 1967.

Eisenberg, L., Lachman, R., Molling, P.A., Locknew, A., Mizelle, J.D. and Conners, C.K.: A psychopharmacologic experiment in a training school for delinquent boys: methods, problems, findings. Am. H. Orthopsychiat., 33, 431-447, 1963.

Farina, A.: Patterns of role dominance and conflict in parents of schizophrenic patients. J. of Abnormal and Social Psych., 61, 31-38, 1960.

Garside, R.F., Birch, H., Scott, D. McI., Chambers, S., Kolvin, I., Tweddle, E.G. and Barber, L.M.: Dimensions of temperament in infant school children. J. Child Psychol. Psychiat., 16, 219-231, 1975.

Gittelman-Klein, R., Klein, D.F., Katz, S., Saraf, K., and Pollack, E.: Comparative effects of methylphenidate and thioridazine in hyperkinetic children: I. Clinical results. Arch. Gen. Psychiat., in press, 1976.

Graham, P., Rutter, M., and George, S.: Temperamental characteristics as predictors of behavior disorders in children. Amer. H. Orthopsychiat., 43, 328-339, 1973.

Gronlund, N.: Sociometry in the classroom. New York, Harper, 1959.

Haley, E.M., Draguns, J.G., and Phillips, L.: Studies of Rorschach content: A review of research literature. Part 2: Non-traditional uses of content indicators. J. of Projective Techniques and Personality Assessment, 31, 3-38, 1967.

Haworth, M.R.: The Children's Apperception Test. New York, Grune and Stratton, 1966.

Henry, W.E.: Projective techniques. In Mussen, P.H. (Ed.): Handbook of Research in Methods in Child Development. New York, John Wiley and Sons, 1960, 603-645.

Hetherington, E.M. and Barclay, M.: Family interaction and psychopathology in children. In Quay, H.C. and Werry, J.S. (Eds.): Psychopathological Disorders of Childhood. New York, John Wiley and Sons, 1972, 30-83.

Kagan, J.: Impulsive and reflective children. In Krumboltz, J.D. (Ed.): Learning and The Educational Process. Chicago, Rand McNally, 1965, 133-161.

Lowenfeld, M.: The world pictures of children. Brit. J. Med. Psychol., 18, 65-101, 1939.

McClelland, D.: Longitudinal trends in the relation of thought to action. J. of Consulting Psychol., 30, 479-483, 1966.

Mensh, I.N., Kantro, M.B., Domke, H.R., Gildea, M.C.L. and Glidewell, J.C.: Children's behavior symptoms and their relationships to school adjustment, sex and social class. J. Social Issues, 15, 8-15, 1959.

Miller, D.R.: Motivation and affect. In Mussen, P.H. (Ed.): Handbook of Research in Methods in Child Development. New York, John Wiley and Sons, 1960, 688-770.

Moreno, J.L.: Sociometry, Experimental Method and the Science of Society. Beacon, New York, Beacon House, 1951.

Morgan, C.D. and Murray, H.A.: A method for investigating fantasies. Arch. Neurol. Psychiat., 34, 289-306, 1935.

Mundy, J.: The use of projective techniques with children. In Wolman, B. (Ed.): Manual of Child Psychopathology. New York, McGraw Hill, 1972, 791-820.

Murstein, B.L.: Theory and Research in Projective Techniques. New York, J. Wiley and Sons, 1963.

Mussen, P.L. and Naylor, H.K.: The relation between overt aggression and fantasy aggression. J. Abn. and Soc. Psychol., 49, 235-240, 1954.

O'Leary, D.: The assessment of psychopathology in children. In Quay, H.C. and Werry, J.S. (Eds.): Psychopathological Disorders of Childhood. New York, John Wiley and Sons, 1972, 234-272.

Patterson, G.R., Ray, R.S., Shaw, D.A., and Cobb, J.A.: Family observation code (adapted from Manual for Coding of Family Interactions). Unpublished ms., 1972.

Rapoport, J., Quinn, P., Bradbard, G., Riddle, K. and Brooks, E.: Imipramine and methylphenidate treatments of hyperactive boys. Arch. Gen. Psychiat., 30, 1974, 789-793.

Rapoport, J.L. and Benoit, M.: The relation of direct home observations to the clinic evaluation of hyperactive school age boys. J. of Child Psychol. and Psychiat., 16, 1975, 141-147.

Richman, N., Stevenson, J.E. and Graham, P.J.: Prevalence of behavior problems in 3-year old children: An epidemiologic study in a London borough. J. Child Psychol. and Psychiat., 16, 1975, 277-289.

Riddle, D. and Rapoport, J.L.: A two year follow-up of 78 hyperactive boys: classroom behavior and peer acceptance. J. of Nerv. and Mental Dis., 1975, in press.

Roff, M.: Childhood social interactions and young adult bad conduct. J. of Abnormal and Soc. Psycho., 63, 1961, 333-337.

Rosensweig, S.E., Fleming, E. and Rosenzweig, L.: The children's form of the Rosenzweig picture frustration study. J. Psychol., 26, 1948, 141-191.

Rutter, M.: Temperamental characteristics in infancy and the later development of behavioral disorders. Brit. J. Psychiat., 110, 651-661, 1964.

Rutter, M. and Graham, P.: The reliability and validity of the psychiatric assessment of the child: I. Interview with the child. Brit. J. of Psychiat., 114, 1968, 563-579.

Rutter, M., Tizard, J. and Whitmore K. (Eds.): Education, Health and Behaviour. New York, John Wiley and Sons, 1970.

Schulman, R.E., Shoemaker, D.H. and Moelis, I.: Laboratory measures of parental behavior. J. Consulting Psychol., 26, 1962, 109-114.

Tchechtelin, M. and Amatora, M.: Children's ratings of associates. J. Exper. Ed., 13, 1944, 20-22.

Thomas, A., Chess, S., Birch, H.: Temperament and Behavior Disorders in Children. New York, NYU Press, 1968.

Ulmer, R.: The Children's Minimal Social Behavior Scale: A short objective measure of personality functioning. Paper presented to the California State Psychological Association, San Diego, Calif., 1967.

Ulmer, R. and Lieberman, M.: Children's minimal social behavior (CMSBS): A short, objective measure of personality functioning (10 year level). Psychol. Rep., 22, 1968, 283-286.

Werkman, S.: The psychiatric diagnostic interview with children. Am. J. Orthopsychiat., 35, 764-771, 1965.

Yarrow, L.: Interviewing children. In Mussen, P.H.: Handbook of Research Methods of Child Development. New York, John Wiley and Sons, 1960.

Yarrow, M., Campbell, J.D., Burton, R.V.: Recollections of Childhood: A study of the Retrospective Method. Monographs of the Society for Research in Child Development. No. 138, Vol. 35, 1970, #5.

Zubin, J., Eron, L.D., and Schumer, F.: An Experimental Approach to Projective Techniques. New York, Wiley, 1965.

APPENDIX X

METHODS OF ENVIRONMENTAL ASSESSMENT*

HOME

Assessing the home environment serves a number of purposes in studies of the effectiveness of psychotropic drugs on psychiatrically disturbed children.

1. It is generally desirable to eliminate from studies children whose home backgrounds are excessively disturbed, first, because for such children there is a question as to whether the child's problems might not be a situational response rather than an underlying disorder; and second, because it is thought that children from good home environments are more likely to show response to treatment. To accomplish this goal, one wants measures of gross psychopathology in the parent and of unusually poor child-raising practices (e.g., child abuse).

2. After children are randomly assigned to treatment or control groups, one needs to demonstrate that family patients are not different between the two groups for two reasons:

   a) If the treated group does significantly better, one wants to be sure this can be attributed to a drug effect, not to the fact that the drug-treated group came from better family backgrounds. The difficulty here is that one does not know along what dimensions one ought to show that family patterns of the drug-treated group are no better. There has been almost no research on family predictors of the course of children's psychiatric disorders (despite a good deal of research on family predictors of the occurrence of disorder in children). Family patterns found beneficial for normal children are not necessarily beneficial for disturbed ones. In normal children, parental permissiveness may be associated with creativity and independence, while studies of factors affecting the course of autism and conduct disorders suggest children with these disorders do well in highly structured, rather rigid environments.

   b) One wants to show that the treatment and control groups include similar proportions with probable genetic vs. socio-cultural etiology, in case this distinction is an important predictor of susceptibility to treatment or to placebo effects. Therefore, to the extent that the childhood disorder is thought to have a genetic component, one would want to evaluate the family history for the presence or absence of that disorder.

3. Once a treatment has been shown statistically to be more effective than placebo, one wants to know the degree to which it bridges the gap between pathological and normal behavior. To do this, one wants to compare treated children with normal controls from similar backgrounds, since one hardly expects the drug to compensate for differences in parents. For this purpose, one wants to measure family characteristics known to be associated in the general population with whatever outcome variables will be used. If outcome measures are to be school difficulties, delinquency, drinking and other such well-studied evidences of childhood deviant behavior, there is a large literature on psychiatric disorder in parents, social status, and family size that can be used. For

*Written by L. Robins

disorders characterized by subjective symptoms rather than behavior problems, little is known about what family characteristics are relevant.

One wants, in matching families of patients and normal controls, to avoid measures of family characteristics based on the quality of the parent's interaction with the index child, since we know that the quality of that interaction will have been affected by the treated child's past and present psychopathology. If normal children were to be matched to treated children with respect to parents' approval of their behavior, for instance, one might have to pick normals whose parents are hypercritical!

4. After a drug has been shown to be effective as compared with placebo, it is valuable to analyze the treated sample with respect to which subgroup showed the most benefit from the drug (and which the least), to inform clinicians about any populations in which it is especially useful. One way of dividing the treated population into subgroups is along family dimensions. These dimensions will be useful to clinicians only if easily recognized by them. Therefore, measures should be very simple ones, such as family welfare status, family size, or mother's education. No special measurement scales are recommended for this purpose, since one could not expect a clinician to use them.

To summarize, the following types of family assessments would be useful:

Measure	Use
Psychopathology in parents	Excluding cases initially (1)
	Assessing proportions with probable genetic factors in treated vs. placebo group (2b)
	Matching treated cases with normals (3)
Quality of child-rearing	Excluding extremely disturbed cases initially
Gross family descriptors: size, mother's education, welfare status	Matching treated cases with normals (3) Describing "good responders" (4)

Measuring gross family descriptors would not seem to require any special instruments. The measurement of family psychopathology and quality of child-rearing does require instruments. Available instruments in the literature have been sought and evaluated according to the following criteria:

Reliability

Validity

Age of children to which appropriate

Ease of administration and scoring

Usefulness in varying socioeconomic settings

Usefulness with informants of both sexes

Independence of quality of interactions between parent and the index child.

Only those instruments were considered which are 1) self-administered questionnaires or 2) interviews which are sufficiently structured so that they can be given by persons without special educational requirements and scored in a routine fashion. These two requirements meant omitting all inventories based on observations of family interaction (e.g., Johnson and Lobitz), open-ended interviewing (Greenberg, Brown and Rutter), and summaries from case records (Geismar).

The review was based on recent publications by Resources in Education, (ERIC), Psychological Abstracts, Abstracts for Social Workers, tests collected by Educational Testing Service, and references in the literature.

Measures of psychopathology in the parents (and extended family):

Recommended: Katz Adjustment Scale (for relative)

This scale is administered to one adult about another. There is a version to be administered directly to the individual, but that form neglects areas of pathology of known importance to children's disorders - alcohol, drugs, and antisocial personality. Designed to be used with former hospital patients, it covers the whole spectrum of serious psychopathology, but has also been shown to be usable in normal populations and to distinguish them from psychiatric patients.

It covers treatment, homemaking, work, family relationships, symptoms, social interaction, alcohol and drug use, and police experience. It is self-administered and easy to score.

There are three major drawbacks: 1) it covers only the last few weeks and will therefore omit disorders in remission, which may have been important earlier in the child's environment or have been passed on to him genetically; 2) it must be asked about a parent rather than to the parent; 3) although it asks all the symptoms necessary for diagnosis, it provides no criteria for making specific diagnoses. Therefore, an overall level of pathology can be obtained but not a meaningful profile of type of pathology. Unfortunately, there appears to be no instrument available which is preferable.

Measures of quality of child-rearing:

There are many more instruments for assessing quality of parenting and family atmosphere than parental psychopathology. However, almost all ask about interactions with a designated index child, which makes them inappropriate for comparing treated and normal children, and none specifies a normal-pathological breakpoint that would be grounds for excluding children from a study on the basis of abnormally poor parenting. Nor do any include questions about the more extreme types of undesirable parent behaviors, such as beatings, long-term isolation, serious food deprivation, cursing, etc. They are all designed with normal populations in mind, and do not address themselves to serious pathology.

Recommended: Moos Family Environment Scale

This scale is impressive in terms of its broad topic coverage (family control, organization, religious activity, achievement interest, intellectual interests, expressive and artistic interests, cohesion, conflict, recreational activities), the fact that any adult member of the family could answer it for the family as a whole, its ease of scoring, its independence of relationships especially involving the index child, and the fact that it has been shown to distinguish a psychiatric clinic sample from a normal population. It has been shown to have fairly good reliability over an eight-week period (subscales correlated between .68 and .86)

but there has been no testing of agreement among members of the same family. It would be appropriate for describing families including children of various ages.

Most other tests considered are contaminated by description of the index child's behavior. This is true of both the Henderson HELPS and the Strom Parent as a Teacher (PAAT) tests, which include items about how the child plays as well as the mother's behavior toward him. Both of these also ignore the father, if there is one. The Block Child-Rearing Practice Report is a well designed Q-sort method but of unknown validity and reliability. (Reliability tests are only on non-parents reporting about their own childhoods.) It also concentrates on relationships between parent and index child, as does McDaniel's Purdue Questionnaire for Parents of Primary School Children. Schaefer and Bell's Parental Attitude Research Instrument (and Dibble's Parent Report developed from it) investigates attitudes rather than behaviors, and results have been found poorly correlated with behavior. The Schaefer and Bell scheme assesses attitudes in terms of degree of control, hostility, and democraticness.

It is thought that certain intersections along these three dimensions are beneficial for children, but these desirable points are not clearly specified. Further, the scale was developed with very young children in mind, and is probably not appropriate after the age of six.

SCHOOL

The problems in assessing home environments are magnified when it comes to finding scales for describing school environments. There is no agreement in the literature that the quality of a school has a measurable effect on academic achievement, much less on psychiatric status. Most attempts to show an effect of school programs on children have been negative. This is true when special programs are evaluated, looking, for instance, for an effect of Headstart on IQ and early reading, or vocational-educational programs on delinquency, or of special schools on autism. It is equally true when attempts are made to show that quality of teachers and excellence of facilities affect academic achievement in ordinary schools (the Colemen Report).

In the absence of knowledge to the contrary, however, it is possible that school environment might have importance in the assessment of drug effects similar to those proposed for family environment: 1) children from extremely noxious school environments might not be ill, but only responding to that environment and therefore should be omitted from the test, 2) treated groups might appear to improve more than controls only because their schools were better, 3) treated children should be compared with controls in equally therapeutic or noxious school settings, and 4) drugs may be more effective in children in some kinds of school settings than in others.

To be useful in studies testing drug effectiveness, evaluations of school environments would have to be based on simply gathered data, not on elaborate observation techniques nor on compilations of administrative data. Like family assessments, questionnaires or interviews seem most practical. But it is not clear who the respondents should be - teachers, students, parents? While this may be an important question, the paucity of appropriate instruments available limits our choices.

Recommended: The Wrightsman School Morale Scale

This scale is designed to be completed by students. It has been given to children over ten years of age. To evaluate schools for younger children, one would have to rely on their older schoolmates' opinions. Whether this would be satisfactory depends on how consistent "school morale" is over various class levels.

The School Morale Scale measures children's views of the quality of buildings, teaching, administration, community opinion of the school, relations among students and between

teachers and students, and the students' feeling about attending. It has been used successfully in schools at various socioeconomic levels.

There is one interesting result -- within a school, children who do well (not surprisingly) rate the school higher than those who do not. This suggests that a far easier way of evaluating schools might be to use the mean and standard deviation of the students on national achievement tests. Of course, it is true that we have no evidence that a school in which a high proportion of students do well is necessarily therapeutic for disturbed children. Schools with high-achieving students must either attract gifted students (because of favorable location or a good reputation), or teach well or both. At this point, we do not know whether either of these characteristics (an environment of gifted students or good teaching) can be shown to help disturbed children, but until the requisite research is done, it would seem as reasonable and as simple to match comparison groups on the general level of achievement of their classmates as on any other school characteristic.

REFERENCES

Block Jeanne H: The Child-Rearing Practices Report. Institute of Human Development, University of California, Berkeley, California, 1965.

Brown GW, Rutter M: The Measurement of Family Activities and Relationships: A Methodological Study. Human Relations 19: 241-263, 1966.

Coleman JS: Equality of Educational Opportunity. National Center for Educational Statistics, U.S. Government Printing Office, Washington, D.C., 1966.

Dibble E, Cohen DJ: Companion instruments for measuring children's competence and parental style. Arch Gen Psychiatry 30(6): 805-815, 1974.

Geismar LL, Ayres B: A Method for Evaluating the Social Functioning of Families under Treatment. Social Work 4(1): 101-108, January 1959.

Greenberg JW: Parent Interview Schedule., in "Home background and school achievement of black urban ghetto children," Am J Orthopsychiatry 42(5): 803-810, 1972.

Henderson RW, Bergan JR, Hunt M: Henderson Environmental Learning Process Scale. Development and Validation of the Henderson Environmental Learning Process Scale, J Social Psychology 88: 185-196, 1972.

Johnson SM, Lobitz GK. Parental manipulation of child behavior in home observations: a methodological concern. J Applied Behavior Analysis 7(1): 23-24, 1974.

Katz Adjustment Scale. Katz MM, Lyerly SB: Methods of measuring adjustment and social behavior in the community: 1. Rationale, description, discriminative validity in scale development. Psychol. Reports 13: 503-535, 1963.

McDaniel ED: Purdue Questionnaire for Parents of Primary School Children. Educational Research Center, Purdue University, Lafayette, Indiana.

Moos RH: Family Environment Scale. Consulting Psychologists Press, Inc., 1974

Schaefer ES, Bell RQ: Development of a Parental Attitude Research Instrument. Child Development 29(3): 339-361, 1958.

Wrightsman LS, Nelson RH, Taranto M: School Morale Scale. George Peabody College, Nashville, Tennessee, 1968.

APPENDIX XI

ASSESSMENT OF SIDE EFFECTS IN CHILDREN*

It is well recognized that side effects of drugs should be studied as carefully and systematically as effects on target behaviors. Surveillance for side effects should include not only emergence of undesirable symptoms but loss of desirable functions as well. In addition, children should be examined for the appearance of side effects both during the time drug treatment is in effect and upon withdrawal of drugs. Where possible, children as well as caretakers should be interviewed directly. Because children often fail to report symptoms spontaneously, care should be taken to inquire specifically about side effects of interest. At present, there are essentially no scales for evaluation of side effects for which reliability has been established. Review of the literature did not reveal any specific checklists devised to record or assess drug-induced side effects in children that have been used extensively or repeatedly. A survey of 20 clinical investigators revealed that most had either assembled scales for a particular medication, or had used the ECDEU Scales (DOTES or STESS) since they became available.

The New York University Unit on Early Clinical Drug Evaluation has used primarily three forms for their studies with young autistic populations (M. Campbell, personal Communications).

These are 1) A lithium toxicity checklist, 2) Haloperiodol-Chlorpromazine side effect checklist, and 3) Toxicology Checklist.

In none of the listed are the items defined in any manner, are instructions given as from whom the information is to be obtained, nor do the listed scales ever give cues or hints, etc., as to how to elicit the information. Furthermore, none of the scales offer anchor points for the various rating levels.

In 1972, Drs. Soltys and DiMascio developed a side effect checklist to be used specifically in drug studies with children. By giving examples of questions to ask the children - or what to observe in them - they gave a rough definition of the symptom items and structured somewhat the interview with the child. In instances where communication with the children was impossible, the child's parents, teachers, nurses, etc., could be used as the informant. All items were to be assessed on a four-point scale (0-none to 3-very much) rather than on a simple present-absent dichotomy. While reliability in using the scale appeared very high in their hands, their limited use precluded any formal testing of the reliability.

The STESS (Children's Self-Rating Treatment Emergent Side Effect Scale) was developed as part of the ECDEU Pediatric Psychopharmacology Battery in 1974. To date, this scale has only been filled out on about 100 children, according to information supplied by the Biometric Laboratory.

All side-effects checklists examined make liberal use of verbalizations from the child in making assessments of adverse effects. Yet there are populations of children in whom these drugs are tried who are incapable of offering any insight into their feelings, reactions, etc.

*By A. DiMascio

A side-effect checklist should be developed in which cues are specified that might reflect adverse effects occuring.

For example

Irritability: reacts negatively--with yelling at or pushing away of individuals attempting to interact.

Restlessness: child paces about, shows constant movement or alters position frequently (sitting to standing up, etc.).

Mouth Dryness: wetting of lips, desire for water

To date this has not been done systematically but is clearly warranted for populations of children such as autistic or retarded.

All checklists described here are attached.

CATALOGUE OF SYMPTOMS - DOTES

NEUROLOGIC:	Akathisia
AUTONOMIC:	Dry Mouth Nasal Congestion Blurred Vision Constipation Increased Salivation Sweating Nausea/Vomiting Diarrhea
CARDIOVASCULAR:	Hypotension Syncope/Dizziness Tachycardia Hypertension EKG Abnormality
OTHER:	Dermatologic Weight Gain Weight Loss Anorexia/Decreased Appetite Headache Tardive Dyskinesia

Rated on a 0 - 4 scale

0 = Not Assessed
1 = Not Present
2 = Mild (Doesn't hinder functioning)
3 = Moderate (Impairing but not hazardous)
4 = Severe (Definite hazard or incapacitation)

Also on a 5 point scale if related to drugs given

POST TREATMENT*

Wt.____ P.____ B.P. ____

Side effects:

appetite decrease	spon.	inq.
appetite increase	spon.	inq.
blurred vision	spon.	inq.
difficulty falling asleep	spon.	inq.
difficulty staying asleep	spon.	inq.
difficulty awakening	spon.	inq.
diplopia	spon.	inq.
drowsiness	spon.	inq.
early awakening	spon.	inq.
dry mouth	spon.	inq.
flushing	spon.	inq.
headaches	spon.	inq.
irritability	spon.	inq.
nausea	spon.	inq.
sadness	spon.	inq.
short breath	spon.	inq.
sweating	spon.	inq.
tremor	spon.	inq.

*This scale was devised by J. Rapoport, M.D., for evaluation of side effects associated with methylphenidate hydrochloride administration. Dr. Rapoport stated that spontaneous reports of side effects negatively correlated to I.Q.

MH-9-36
1-73

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
HEALTH SERVICES AND MENTAL HEALTH ADMINISTRATION
NATIONAL INSTITUTE OF MENTAL HEALTH

FORM APPROVED
OMB NO. 68-R66

STESS

PATIENT INITIALS

FIRST INITIAL

SECOND INITIAL

FORM NO.

NUMBER MALES 001 TO 499; FEMALES 500 TO 998

PATIENT

RATER

S
M
F

PERIOD

Hours
Days
Weeks
Months

PLEASE USE A NO. 2 LEAD PENCIL. BE SURE TO MAKE MARKS HEAVY AND DARK. ERASE COMPLETELY ANY MARKS YOU WISH TO CHANGE.

INSTRUCTIONS: Since the last time here, have you been bothered with or had trouble with any of the items listed below? If this is your first visit, have you been bothered by any of these items in the last week? Mark the number which best tells how much you were bothered. When filling out form for the child, mark on the basis of what you have seen or what the child has complained about. If you are unsure, mark "Don't know".

EXAMPLE	Not at All	Just a Little	Pretty Much	Very Much	Don't Know
Cramps?	0	1	2	3	4

Have you had trouble with:

ITEM	Not at All	Just a Little	Pretty Much	Very Much	Don't Know
1. Eating?	0	1	2	3	4
2. Drinking?	0	1	2	3	4
3. Dry mouth and lips?	0	1	2	3	4
4. Wetness in mouth?	0	1	2	3	4
5. Fewer bowel movements (constipation)?	0	1	2	3	4
6. More bowel movements (diarrhea)?	0	1	2	3	4
7. Stomach aches?	0	1	2	3	4
8. Muscle cramps?	0	1	2	3	4
9. Being sick to your stomach?	0	1	2	3	4
10. Wetting the bed?	0	1	2	3	4
11. Urinating?	0	1	2	3	4
12. Itchy or scratchy skin?	0	1	2	3	4
13. Rashes?	0	1	2	3	4
14. Colds or sniffles?	0	1	2	3	4
15. Headache?	0	1	2	3	4
16. Dizziness?	0	1	2	3	4
17. Playing sports?	0	1	2	3	4
18. Shakiness?	0	1	2	3	4
19. Pronouncing words?	0	1	2	3	4
20. Doing things with your hands?	0	1	2	3	4
21. Sitting still?	0	1	2	3	4
22. Tiredness?	0	1	2	3	4
23. Feeling sleepy?	0	1	2	3	4
24. Trouble getting or staying asleep?	0	1	2	3	4
25. Bad dreams?	0	1	2	3	4
26. Getting along with parents?	0	1	2	3	4
27. Getting along with other kids?	0	1	2	3	4
28. Crying?	0	1	2	3	4
29. Getting mad?	0	1	2	3	4
30. Not being happy?	0	1	2	3	4
31. Being sad?	0	1	2	3	4
32. Paying attention?	0	1	2	3	4

LITHIUM TOXICITY CHECK LIST*

Gastrointestinal symptoms

1. Anorexia
2. Nausea
3. Vomiting
4. Diarrhea
5. Constipation
6. Dryness of the mouth
7. Metallic taste

Neuromuscular symptoms and signs

1. General muscle weakness
2. Ataxia
3. Tremor
4. Muscle hyperirritability
   a. Fasciculation (increased by tapping muscle)
   b. Twitching (especially of facial muscles)
   c. Clonic movements of whole limbs
5. Choreoathetotic movements
6. Hyperactive deep tendon reflexes

Central nervous system

1. Anesthesia of skin
2. Incontinence of urine and feces
3. Slurred speech
4. Blurring of vision
5. Dizziness
6. Vertigo
7. Epileptiform seizures
8. Electroencephalographic (EEG) changes

Mental symptoms

1. Difficulty concentrating
2. Slowing of thought
3. Confusion
4. Somnolence
5. Restlenessness—disturbed behavior
6. Stupor
7. Coma

Cardiovascular system

1. Pulse irregularities
2. Fall in blood pressure
3. Electrocardiographic (EKG) changes
4. Peripheral circulatory failure
5. Circulatory collapse

LITHIUM TOXICITY CHECK LIST (Cont'd.)

Miscellaneous

1. Polyuria
2. Polydypsia
3. Glycosuria
4. General fatigue
5. Lethargy and a tendency to sleep (drowsiness)
6. Dehydration
7. Skin rash--dermatitic lesions
8. Weight Loss
9. Weight gain
10. Alopecia
11. Quincke's edema

*Checklist prepared by B. Shopsin and S. Gershon (1973).

HALOPERIDOL- CHLORPIOMAZINE SIDE EFFECT CHECKLIST

ANOREXIA

LOSS OF WEIGHT

WEIGHT GAIN

SEDATION

(WORSENING OF) HYPERACTIVITY

(WORSENING OF) HYPOACTIVITY

IRRITABILITY

POLYURIA

DROOLING

DYSTONIA

AKATHISIA

TREMOR

DRY MOUTH

TROUBLE URINATING

CONSTIPATION

COGWHEEL PHENOMENON

TOXICOLOGICAL RATING: Children's Psychopharmacology Unit
N.Y.U. Medical Center

Patient _____ Hospital No. _____ Study No. _____

Date: _____ Drug: _____ Dose: _____														
Temperature														
Pulse														
Respiration														
BP														
Weight														
Pupils D/M/C														
"Pasty"														
Mottling														
Flushing														
Itching														
Skin Rash														
Jaundice														
URI														
Dry Mouth														
Stomach-ache														
Nausea														
Vomiting														
Anorexia														
Diarrhea														
Constipation														
Headache														
"Dizzy"														
Malaise														
Lethargy														
Drowsiness														
"Sleepy" (Subj)														
"Sleepy" (Obj)														
Irritability														
Agitation														
Insomnia														
Restless														
Unsteady Gait														
Dystonia														
Rigidity														
Tremors														
Seizures														
Masked Facies														
Other														

TOXICOLOGICAL RATING SHEET #2 (2-6 year patterns)

Name _____

Birthdate _____

Hospital No. _____

Study No. _____

Date _____																
Drug _____																
Dose _____																
Sleeping - Night																
1. Hour to bed																
2. Hour to sleep																
3. Wakefulness *																
a. Hour awake																
b. Hour to sleep																
c. Restlessness *																
d. Irritability *																
4. Hour awakened - AM																
a. Awoke spontaneously																
b. Had to be awakened																
Sleeping - Nap																
5. Hour to bed																
6. Hour to sleep																
7. Hour awakened																
a. Irritability *																
b. Lethargy *																
c. Awoke spontaneously																
d. Had to be awakened																
Special Symptoms																
1. Head banging *																
2. Rocking *																
Eating - Lunch **																

* 0 = none; sl = slight; m = moderate; s = severe

** 0 = none; + = poor; ++ = fair; +++ = good

CHILDREN'S SIDE-EFFECT LIST

	Study No.	Patient No.	Period	Form No.	Drug Code
Enter Nos					
Col. No.	1	2 3 4	5 6	7 8	9

Name of Child_____ Age_____ Date_____ Grade_____

Weight_____ School_____ No. capsules daily

Other Medication_____ (code) 13-31

(To be completed by a project nurse and based on reports of 1. child, 2. parent, 3. teacher, 4. physician, 5. nurse. Whenever contradictory answers are given by child and adults, note item of contradiction and use all resources to come up with a judgement.)

Interval History (for the period since the last "side-effects" list was completed) Ask the child if anything new has happened to him or the family and note any significant illness, changed routine or environment).

Instructions for Side-effects Questionnaire

1) Specify, when questioning the child or others, that the questions asked pertain only to experience since the last form was completed.
2) Questions or descriptions in parentheses should be used as guides in eliciting information directly from the child relative to the incidence and degree of the symptom.

General

(To help the child focus on what, if anything, is disturbing him in relation to the medication, such questions as the following might help elicit spontaneous comments.)

1. Did you notice anything different about yourself since the last time I asked you about the medicine?
2. Does the medicine change you?
3. How does it make you feel different?
4. Do you want to take it? Why? Why not?

(Each specific item must be checked)

RATINGS: 0 = None
1 = A Little
2 = Moderate
3 = Very Much

I. CENTRAL NERVOUS SYSTEM

1. Drowsiness (Do you have trouble keeping awake in school? Other places? Have you fallen asleep at school? _____ 32
2. Headache (Have you had any headaches? How bad? Where – point – did it hurt? What brought it on? What made it better?) _____ 33
3. Irritability (Is it harder for you to keep your temper? Is it harder to get along with your family? Other kids? Do you get angry more easily or more often? _____ 34
4. Sensitivity (Do you cry more easily or more often? Do things that didn't use to make you feel sad now make you feel like crying? Do you cry for no reason at all?) _____ 35
5. Restlessness (Is it harder for you to sit still? Do you feel more like getting up, moving around?) _____ 36
6. Tremor (Do your fingers or hands shake when you try to do things with them? Do they shake so hard you can't do certain things? Show me.) _____ 37
7. Incoordination (Have you had trouble walking or keeping your balance? Any trouble playing sports or games like baseball or jumping rope? What kind of trouble) _____ 38
8. Insomnia (Do you have trouble getting to sleep at night? What keeps you awake? Do you wake up alot during the night? Why?) _____ 39
9. Speech Impairment (Speech is slurred or monotonous in tone. Is it harder to say words? Why?) _____ 40
10. Seizures (Record time, place, circumstances and detailed description, e.g., whether child fell, bit tongue, wet self.) _____ 41

_____

_____

_____

II. DYSTONIC SYMPTOMS

1. Muscle Spasms/Twitches
   a) Lips, Tongue (Do you have trouble talking? Does your tongue keep moving?) _____ 42
   b) Swallowing (Drooling, difficulty drinking or swallowing food.) _____ 43
   c) Others (Specify: e.g., oculogyric crises, opisthotonus.) _____ 44

_____

_____

2. Muscle Rigidity (Observe directly, also ask if muscles feel stiff and hard to move.) _____ 45
3. Atonia (Loss of muscle tone – observe directly – also ask if muscles feel weak and no strength to move them.) _____ 46

III. CARDIOVASCULAR

1. Dizziness (How do you feel when you first wake up or just after you spin around fast? Do you feel weak in the legs and things around you look dark and fuzzy?) _____ 47
2. Fainting (Did everything turn black and you couldn't remember what happened? Did you fall? Did anyone see what happened?) _____ 48
3. Palpitations (Does your heart feel like it is beating too fast? Too hard? Too slow?) _____ 49
4. Chest Pain (Do you have pains in your chest – pointing to it –? What is the pain like? What brings it on? Makes it better?) _____ 50
5. Breathing Difficulty (Is it hard to catch your breath? Does it feel like something is squeezing your chest? Do you feel like you're not getting enough air?) _____ 51
6. Nosebleed (Have you had nosebleeds? How many? What brought it on? _____ 52
7. Nasal Congestion (Does your nose feel stuffy? Has it been "running"? Does it feel like you have sniffles?) _____ 53

(With moderate and severe responses to items in this section, check and record)

Blood Pressure_____

Respiration Rate_____

Pulse Rate_____

Describe any unusual features in respiration such as difficulty, labored, sighing, gasping, etc.

_____

_____

_____

IV. GASTROINTESTINAL

1. Increased Appetite (Do you feel like eating all the time? Do you eat more?) _____ 54
2. Decreased Appetite (Do you skip any meals now? Do you finish lunch? Do you feel less like eating between meals?) _____ 55
3. Abdominal Pain (Does your tummy hurt? Where? When? What makes it better?) _____ 56
4. Nausea (Do you feel like throwing up? When? What makes it feel better? _____ 57
5. Vomiting (Do you throw up? When? How much? What do you bring up? What makes it stop?) _____ 58
6. Mouth Dryness (Does your mouth feel dry all the time? Anytime? Real dry or just a little?) _____ 59
7. Constipation (Do you have trouble having bowel movements? How many do you have each day? How many now? Do you notice what they look like? Any blood?) _____ 60
8. Diarrhea (Do you have trouble having too many bowel movements or the "runs"? Are they hard to control?) _____ 61
9. Miscellaneous (Specify, e.g. bad taste, increased salivation.) _____ 62

_____

_____

_____

V. GENITO-URINARY

1.	Urinary Frequency	(Do you have to urinate ("p", "tinkle") more now? How many times while in school? Do you have to get up during the night?)	63
2.	Urinary Retention	(Is it harder for you to urinate ("p", "tinkle") now?	64
3.	Painful Urination	(Does it hurt when you urinate ("p", "tinkle") now?	65
4.	Bedwetting	(Do you sometimes wet your bed at night? Every night? How often?)	66
5.	Miscellaneous	(Specify, e.g., daytime incontinence?)	67

VI. EYES

1.	Blurred Vision	(Do your eyes see things "fuzzy"? Do you have trouble reading because letters are not clear?)	68
2.	Light Sensitivity	(Do your eyes hurt from lights inside the school or at home? How about sunlight?)	69
3.	Itching/Tearing	(Do your eyes hurt, burn, itch, make alot of tears? Observe the conjunctiva.)	70
4.	Bloodshot – Jaundiced	(Observe the sclera)	71
5.	Nystagmus	(Specify the type observed, e.g. laterality, direction, frequency.)	72

VII. SKIN

1.	Itching	(Do you itch anywhere? Where? What brings it on? What makes it stop? Scratching?)	73
2.	Rash	(Do you have red spots anywhere on your skin? Describe if rash is present.)	74
3.	Skin Color	(Specify, e.g., pallor, flushing, sweating, coldness, mottling, pigmentation, jaundice.)	75
4.	Purpura	(Have you noticed any bleeding under the skin? Do you get "black and blue" spots easily? Do you have trouble stopping bleeding from cuts?)	76

Prepared by:

John Soltys, M.D.
Alberto DiMascio, Ph.D.
Boston State Hospital
591 Morton Street
Boston, Mass. 02124

☆ U.S. GOVERNMENT PRINTING OFFICE : 1979 O—295-980