Drug Screening and Evaluative Procedures

Current approaches do not provide the information needed for properly predicting drug effects in man.

Samuel Irwin

". . . observation is the mind's support in reasoning, and experience the mind's support in deciding . . ." (*1*).

The pharmacologist engaged in the screening and evaluation of drugs is required to select drugs of potential interest from among thousands of new chemicals presented for testing. According to Smith (*2*) and de Haen (*3*), out of 114,600 different chemical substances tested by the pharmaceutical industry in 1958, approximately 1900 were selected for clinical trial. Of these, only 44 (2.3 percent) were ultimately marketed—a figure that indicates an inordinately high attrition rate. Since the cost of a single drug evaluation, with assessment of long-term toxicity and minimal clinical effect, may be tens of thousands of dollars, this attrition rate is of no small consequence. It leads one to examine the extent to which failures in selecting drugs and in predicting their usefulness are unavoidable at our present state of knowledge, or may arise from the manner in which we accumulate information and evaluate drugs.

The action of some drugs, such as acetylsalicylic acid or phenylbutazone, is demonstrable only in man. For agents of this type, laboratory tests of activity at present are of little or no predictive value. In the case of other drugs, however, where one can demonstrate similar activity in animals and man, one can proceed from animal studies to human studies with greater assurance. Nonetheless, some investigators place little reliance on findings in animals, particularly with respect to drugs that affect behavior, and emphasize instead the importance of clinical screening and evaluation of almost any drug found in animal studies to possess interesting or unusual activity. They point to the fact that almost all unique drug actions have been discovered in studies in man. Other investigators, however, feel that there exists a large area of direct or empirical observation in animals from which clinical predictions can be made with reasonable assurance. These investigators place greater reliance on the results of testing and drug evaluation in animals and consider it possible, on the basis of animal studies, to discard many of the drugs that might otherwise be submitted for testing in man. As a consequence, they submit fewer drugs for clinical trial.

Thus, much of what is done in the laboratory today is predicated on a variety of biases and attitudes. Some of these attitudes have arisen from the development of pharmacology as an experimental science, for investigators have been led to adopt methods and experimental approaches in drug screening and evaluation which, in important respects, greatly compromise their ability to single out and properly evaluate potentially useful compounds. A major problem, as I view it, is the elusive one of attitude and orientation, since certain seemingly obvious requisites in screening and evaluation seem to be generally overlooked or ignored. Attitude and orientation ultimately determine what the investigator does.

This article represents an attempt to review the problem critically—to describe and analyze some of the more basic requirements for screening and evaluating drugs and some major weaknesses in the procedures as they are carried out today. The article is predominantly oriented toward neuro- and psychopharmacology but is intended as a frame of reference for drug screening and evaluation as a whole.

The author is head of the department of neuropharmacology, Schering Corporation, Bloomfield, N.J.

Screening-Evaluative Program

Drug screening is essentially a scanning procedure designed to distinguish useful from nonuseful drugs as rapidly, comprehensively, and inexpensively as possible. The best approach is one in which no assumptions are made about what the probable action or actions of a compound may be, except when one is dealing with a well-worked-out structure-activity series. In view of the wide variety of measurable drug actions, however, this is a staggering undertaking. In order to screen effectively, the investigator is forced to impose numerous limitations as to the species, route of administration, or procedures employed, and as a consequence he must accept the possibility that he will overlook a useful drug. The task, simply stated, is to develop a program sufficiently comprehensive and balanced to minimize this possibility.

One of the most important aspects of drug screening and evaluation is the problem of decision making—the "yes" or "no" with respect to interest in a compound. It is important to obtain in the earliest phase of screening the information essential to making this decision, so that compounds of little or no apparent value can be discarded quickly. To accomplish this, it seems best to proceed from the general observation to the specific, and from comprehensive observational techniques, wherever they are applicable, to the use of instrumentation.

This approach accelerates the program and decision making and minimizes the possibility of overlooking unforeseen drug actions of potential clinical usefulness. Instrumental techniques are best brought into play when the time and effort expended in accumulating unit information can be justified, as when they provide the most satisfactory approach to obtaining the needed information or to the further investigation of drugs of established interest.

If it were only a problem of finding a drug with activity, the development of new drugs would be relatively simple. The investigator, however, still faces the task of evaluating the specific action or actions of interest within the context of the overall changes produced by a drug. This requires a considerable degree of sophistication. Ideally, the pharmacologist should be able to predict, within reasonable limits, the effective dosage, actions, side effects, and potential toxicity, as well as the patient-population most likely to benefit from

the drug. In order to even partially achieve this goal, the investigator's prime concern must be the validity, reliability, and relevance of his methods and observations. It is important that he note the whole range of qualitative changes produced by a drug and the quantitative or semiquantitative relationships between them, whether or not these changes and relationships can be accurately measured. He cannot ignore or deny the existence of phenomena merely because they are difficult to measure. In the evaluation of drugs for clinical use, precision of measurement as such is less important than is the accumulation of general information. Moreover, the low portion of the dose-response curve is likely to contain the most germane information.

It is a fundamental concept in drug screening that each drug action involves alterations in the tissue which are reflected in functional changes. These functional changes can be measured, and the data can be used empirically to differentiate and to classify new drugs. To develop relevant, highly specific screens it is desirable, but by no means necessary, to understand the underlying mechanisms of action. Since we generally lack this information, an effort is made to test drugs against the disease process to be treated, or against a suitable model.

A second important concept is that drugs may act similarly in animals and humans, and that laboratory animals can thus serve as "model analogues" of man. However, all biologists are aware that this relationship is tenuous and that vast differences can exist, not only between humans and animals but also between various animal species. It is generally accepted in pharmacology that interspecies differences in response to drugs are least likely to occur in closely related phylogenetic forms, although there are many exceptions to this rule. Generally, however, the probability of carry-over of drug effects from animals to man seems to be greatest when diverse species show similar responses to a given drug, and least when their responses vary widely. It is important, therefore, that new drugs be evaluated for their effects in several animal species, particularly in species from different orders. The range of variation in the responses obtained enables the pharmacologist to establish reasonable limits for error in his predictions.

A third important concept in screening is that drugs are more readily differentiated and classified from multiple measures of their activity—from the overall profile of their action—than from any single measure (*4*). This is true because specific actions frequently arise through different mechanisms and may be shared by different drugs. For example, Houde *et al.* (*5*) have demonstrated in spinal dogs that, although barbiturates, muscle relaxants, and narcotic analgesics all depress the ipsilateral flexor reflex, the drugs can be differentiated and classified if one also considers their differential effects on two other reflexes, the knee jerk and the ipsilateral extensor thrust. As a consequence, the most efficient primary screen would seem to be a comprehensive battery of tests which reveal the dose-response profile, the toxicity, and the pattern-specificity of action of a new drug—a multidimensional approach.

Hazards in Drug Screening

A simple problem of logistics and cost limits much of our screening to studies in small animals, such as the mouse and rat, and imposes a high risk factor, since no single species is a satisfactory indicator of all drug actions. Similarly, use of a single route of administration imposes a high risk factor. The route should be selected with care, for it influences the rate of absorption, the steepness and variability of the dose-response slope, the intensity and duration of the effect, and even the actual profile of the drug action. Some drugs, such as 1-alpha-acetylmethadol, are more active when administered orally than when administered parenterally (*6*).

Selection of an arbitrary single dose of drug or of a specific time of measurement likewise imposes an element of risk. A drug action may be missed altogether, or the actions of low and high doses of the drug may differ qualitatively and mask one another, as in the case of the pressor and depressor effects of epinephrine. Also, the different actions of an agent may exhibit different time-response curves. This is evident in the cat after the administration of chloral hydrate, where one observes peak motor incoordination as an early effect and peak hypnosis very much later.

From a statistical point of view, aside from the dangers implicit in poor experimental design (*7*), special problems arise when one studies the effects of drugs on psychological behavior, for behavioral responses to drugs exhibit much greater variability than do autonomic or neurologic responses. For example, in studying the effects of drugs on locomoter activity or emotional behavior, one may observe diametrically opposite effects in different individuals. This is a not uncommon observation in subjects that have been given phenothiazine tranquilizers; in cats or dogs that have been given these tranquilizers one occasionally observes amphetamine-like stimulation or even increased aggressiveness (*8*). Such findings present special problems in data analysis, for drug effects that are opposite in character may cancel one another out if inadvertently averaged.

Social-interaction effects are another source of difficulty. These effects are sufficiently great at times to obscure the quantitative differences in response of saline- and drug-treated animals (*9*), or to increase the apparent toxic effect of drugs (*10*). Qualitative changes in the response to drugs may also result. For example, doses of phenobarbital or pentobarbital which produce full hypnosis in isolated mice produce marked stimulation when the animals are grouped (*11*). Similar qualitative differences in the response to drugs have been observed in humans (*12*).

As long as we do not have suitable analogues of human disease states when studying drug effects in animals, it will remain impossible for the pharmacologist to predict, a priori, the effect of drugs on such states. It is not surprising, therefore, that the antipsychotic effects of chlorpromazine and the antidepressant effects of monoamineoxidase inhibitors and imipramine were first discovered in man.

Once these effects had been discovered, the pharmacologist could establish a host of laboratory measures which could be used to reveal agents with similar activity. However, although much of what we measure with these agents seems relevant to the treatment of certain behavioral disturbances, such as hyperactivity, agitation, or aggressiveness, there is nothing in their nature which could logically have led us to predict their antipsychotic or antidepressant activity. What is surprising is that a drug like chlorpromazine was not initially recognized as a potentially useful and unique behavioral drug. An analysis of the major weaknesses in drug screening and evaluation, which seem to contribute to present difficulties in drug selection and clinical prediction, is presented below.

Standardized Multidimensional Observation

The evaluation of a patient or a drug requires examination and differential diagnosis. The physician relies heavily on subjective observation, the pharmacologist on special in vitro and in vivo laboratory procedures. The two approaches—those of subjective observation and objective recording—are complementary and furnish different kinds of information. The physician views the total patient; notes significant behavioral, neurologic, or autonomic changes; integrates the information; and often can complete a diagnosis in minutes. The pharmacologist performs a series of laboratory tests, at different times, under different conditions, in different species, and by different routes of administration; integrates the information; and may complete his analysis in weeks or months. The physician observes symptomatic changes within a direct, clinical context; the pharmacologist more often concerns himself with measures and preparations far removed from the clinical situation. Nevertheless, both the physician and the pharmacologist may make an incorrect diagnosis because of excessive reliance on one or the other approach. The physician has turned more and more to the laboratory for the establishment or confirmation of a diagnosis; conversely, the pharmacologist may benefit from a more clinical approach in his evaluations. Both sources of information are required for drug evaluation, and both should be used, due attention being paid to the differences in the nature of the derived information.

Perhaps the greatest deterrent to effective drug screening and evaluation has been failure to give serious consideration to the development of gross observation of animal behavior as a quantitative instrument. Notwithstanding current prejudices as to the scientific status of subjective reporting, a standardized, carefully defined procedure may provide quantitative data of as much reliability and reproducibility as some objective techniques (*13*). Whether labeled "objective" or "subjective," all quantification ultimately requires subjective discrimination. The labels themselves are quite misleading. What is really meant is that some distinctions are more difficult to make than others. From the standpoint of drug evaluation, however, one cannot ignore the existence of discrete measures or constructs, such as mood, malaise, fear, stupor, and apprehension, merely because they are difficult to quantify or too complex to adequately define. Notwithstanding these difficulties, we make such distinctions daily, talk about such constructs, and are somehow understood. We distinguish sickness from health, panic from calm, aggressiveness from friendliness, and we would fare very badly as social animals if we did not. The same principle applies to the study of drug effects in animals. Those engaged in screening and evaluating drugs should carefully consider what the physician has been able to achieve with the gross, even unstandardized and poorly quantified, observation of total behavior as his primary tool.

Standardized, multidimensional observation is a *method* applicable to many experimental situations. It is not a universal panacea; there is much it obviously cannot do. However, it may be the most effective single approach we can use in drug screening and evaluation, particularly in areas of neuro- and psychopharmacology. The major advantage of this technique is that, freed from the restrictions imposed by the use of physical equipment, one can obtain a wide range of quantitative (or semiquantitative), useful, and relevant information about a drug, quickly and economically. Further, much of the information obtained would be difficult or impossible to derive with instrumental techniques.

The use of observational techniques has several major weaknesses, but these are not intrinsic and they are not insurmountable. It would be pointless to deny that individuals differ in their ability to observe or to rate. Observation requires greater attentiveness and skill on the part of the investigator than most other techniques do, and it requires a commensurate degree of training. Thus, in order that differences in findings due to differences in the observational skill of the observers may be minimal, the procedures and rating scales developed must depend as little as possible on skill and experience. This is best achieved where it is possible to describe the behavior in terms of its duration or frequency of occurrence per unit of time or trial or, where intensity is to be measured, by describing and quantifying the observed behavior on the basis of an all-or-none rating scale of events. Thus, for example, the depth of ether anesthesia has been operationally described on the basis of a rating scale of eight progressive steps. It is similarly possible to establish techniques for grading the degree of impairment of the righting reflex, the struggle behavior of animals placed in a series of unusual postures, a visual placing reaction, and many other attributes of behavior or performance (*13, 14*). Once such techniques have been established, observer and interobserver reliability in their application can be readily subjected to statistical analysis and validation.

Certain measures, however, such as malaise, stupor, or fearfulness, are more difficult to define and quantify and will remain a challenge for some time. For measures such as these, which involve a complexity of attributes difficult to describe and quantify verbally, a sequence of pictures or a film which can serve as a standard of reference for describing and scoring the intensity of change would seem the best approach. The point I am making, however, is that rigid criteria such as we apply for the acceptance of objective techniques of measurement can also be applied to observational techniques. The setting up of such criteria merely awaits serious consideration and development.

Other weaknesses in the use of observational techniques are a tendency toward subjective bias, the possibility that the rating scales used may be nonlinear and therefore nonadditive, and a tendency to have an insufficient number of steps in the rating scale. In the latter case, one tends to lose information, whereas excessive subdivision does little harm. How many steps are appropriate in a rating scale, however, and whether the scale is linear or nonlinear can be determined through statistical analysis of the acquired data, and subjective bias can be minimized through the use of blind procedures. The weaknesses attributed to observation, thus, are not intrinsic. They arise in the main from the use of procedures which have not been standardized, validated, and systematically carried out, and the objections would apply equally to objective procedures carried out in similar fashion. If as much time had been devoted to the development of observational techniques as these techniques deserve, there would be available today a variety of standardized procedures for use in animals, producing information of more immediate value and relevance to drug screening, evaluation, and research than is provided by many of our attempts to avoid such direct measurement.

Data Collection and Integration

In drug evaluation it is important to note the whole range of qualitative changes produced by a drug and the quantitative relationships between them. It is unlikely that a drug can be properly evaluated until most of the major tests have been performed, under conditions that are similar, in a single animal species and by the route of administration intended to be used clinically. Such uniformity in testing greatly simplifies the problem of integrating and subsequently interpreting the data. The procedures in use, however, rarely approximate this. More often one finds a tendency to confuse matters by extending the range of variables—by mixing up as many different species, preparations, conditions, and routes of administration as possible in investigating the different actions of a drug. In this way, an enormous mass of data is accumulated which it is almost impossible to integrate. The value of obtaining multiple data from the same species, preferably from the same animals, cannot be overemphasized. It is here that multidimensional procedures are of particular value, for they permit the investigator to obtain a wide range of data from each animal simultaneously and in integrated form. From such data the dose-response relationships for different drug actions can be more meaningfully compared, related, and extrapolated to man.

Dosage

It is generally believed that humans are more sensitive to drug effects than laboratory animals and thus require greatly reduced dosage. This assumption, particularly in neuro- and psychopharmacology, probably derives from the fact that in most evaluative procedures with animals, the responses measured—for example, marked motor incoordination, locomotor stimulation, depression, or elevation of the pain threshold—are quantitatively and often qualitatively far removed from the effects sought clinically. When more sensitive techniques are employed (*13*, *15*) to measure in animals the same indices of change that are observed in humans, we find that an unusual degree of correlation exists between humans and such species as the cat and dog. As a consequence, we have been able to make reasonably valid predictions for man with respect to dosage and the therapeutic ratio of drugs almost entirely on the basis of direct carry-over of such animal findings to man. However, this is possible only when similar events are measured in animal studies and human studies—when emphasis is placed on the therapeutic rather than the toxic effects of drugs.

Such predictions, of course, cannot always be made; species differences in response to drugs do exist. However, when the sensitivity and therapeutic relevance of test procedures is increased, one is likely to find better interspecies correlation with respect to dosage. Where quantitative species differences of response do occur, a linear relationship, across species, frequently exists between the log of the dose administered and the log of the body weight (*16*). When this is true, and when the slope and intercept for a given drug response are known, one can extrapolate dosage to man or to other species with far greater precision than is otherwise attainable.

Quantal versus Graded Measurement

An all-or-none (quantal) approach in measurement greatly simplifies the process of data accumulation and usually facilitates analysis, especially when one is dealing with graded measures such as ataxia or muscle weakness, which are difficult to quantify reliably. This approach is particularly useful in studying relative potency, or in demonstrating the occurrence of a statistically significant change. In the framework of requirements for drug appraisal and prediction, however, all-or-none data can be grossly misleading, of limited value, and at best a poor substitute for graded, quantitative information when the end points selected for analysis are *extreme* drug-induced changes. Such data place the investigator in the uncomfortable position of having to predict the effective dose, therapeutic ratio, or side effects of a drug in man from almost irrelevant information. In drug screening, a quantal approach imposes the additional danger that the investigator will overlook potentially useful and safer drugs which may be unable to produce the marked changes required by the procedure in use.

All that is actually demanded of a drug is that it produce the degree of quantitative change desired in man, yet many laboratory criteria for drug activity greatly exceed this level. The value of an all-or-none approach would seem entirely contingent upon one's objective; this should be clearly defined. In drug screening and evaluation, precision of measurement, as such, is far less important than is the accumulation of meaningful information.

Use of Indirect Measures

There appears to be a tendency to employ indirect measures for drug actions which can be measured directly, the principle hazard being that one may not measure what one intends to. It makes no sense at all, for example, to measure abolition of the righting reflex ("sleep time") as an index of hypnotic activity. This reflex has nothing to do with sleep. Empirically, many hypnotic drugs do in fact abolish this reflex, but so do muscle relaxants, neuromuscular blockers, narcotic analgesics, tranquilizers, and many other classes of drugs. Depending upon the frame of reference, the same response may be labeled hypnotic dose (HD_{50}), anesthetic dose (AD_{50}), or paralytic dose (PD_{50}). Even worse, most investigators using this measure do not differentiate between the behavioral (level-of-awareness) and neurologic (neuromuscular) abolition of the righting reflex, a differentiation easily made by applying pressure or a pain stimulus to the animal's tail. Further, by taking motor impairment as an end point, one builds this property into every hypnotic or muscle relaxant developed. Similarly, if one uses spinal-reflex changes like those produced by mephenesin as a screen for muscle relaxants, one is likely to select mephenesin-like muscle relaxants only, and to effectively eliminate muscle relaxants which produce their effects through different mechanisms, as does chlorpromazine after supra-tranquilizing doses.

The potentiation of barbiturate anesthesia is another indirect measure widely used in pharmacology. While there can be no objection to using the procedure to measure potentiation of or antagonism to barbiturate activity, all too frequently one finds it used as a measure of hypnotic activity. Nothing could be more misleading. According to Riley and Spinks (*17*), "20 to 30 per cent of randomly selected compounds are able to prolong hexobarbitone sleep when given (by pretreatment) in the relatively modest dose of 100 mg/kg,

and their subsequent examination is a formidable task."

The point is that changes in the level of wakefulness or of skeletal-muscle activity can be measured directly, without resort to inappropriately labeled procedures which bear little relation to what they are supposed to measure. The pharmacologist, it would seem, should devise methods for making more direct measurement and, more important, should address himself more completely to the question of what is being measured and what the results are likely to mean.

Use of Intact Animals

In studying the effects of drugs, except where diseased states or abnormal animal preparations are required to demonstrate certain drug actions (for example, anti-inflammatory or anticonvulsant activity) one should avoid the pitfalls of relying on other than intact animals that have received no medication. In biology, the whole is more than the sum of its parts. Analysis to ever-finer levels of structure does not necessarily reveal the nature of relations between the parts and may actually destroy the relations. Although the dynamics and mechanisms of drug action are most readily studied in isolated situations, the reliability of knowledge as a basis for clinical prediction tends to decrease as we proceed from the whole to isolated tissue, single functional units, or subcellular levels of activity.

Some of the dangers in evaluating drugs solely at the enzyme, electrophysiological, or neurophysiological level have been emphasized recently by Toman (*18*), Bain (*19*), and Killam and Killam (*20*). These authors note that surgical lesions, in vitro administration of drugs, or the use of anesthetics or neuromuscular blocking agents frequently affect responses to drugs, causing them to differ greatly from responses obtained in normal animals. Thus, although it is sometimes more convenient or expedient to employ unusual or isolated preparations, an intact animal that has received no medication must be used wherever possible in evaluating a drug's activity. As Kety aptly states it (*21*), "we do not always get closer to the truth as we slice and homogenize and isolate," what we gain in precision and in the rigorous control of variables we sometimes lose in relevance to normal function, and, in the case of certain diseases or problems, the fundamental process may often be lost in the cutting.

Long-term versus Short-term Study and Individual versus Group Study

Another area requiring emphasis is the almost universal failure to study drug effects under long-term as well as short-term conditions of drug administration. The pharmacologist is required both to find useful drugs and to predict their clinical efficacy. In screening new agents on a single-dose basis one cannot recognize as useful a drug whose actions are demonstrable only after long-term administration. Similarly, for drugs of established interest likely to be administered repeatedly, it seems important to determine whether tolerance, increased sensitivity, or cumulative effects develop in response to any or all of its actions. The end result of long-term administration of a drug may be more revealing and significant than the short-term effects (*4*). In this connection, one should perhaps take advantage of long-term toxicity studies to investigate possible withdrawal effects of drugs before the animals are sacrificed.

In the analysis of data, altogether too much emphasis has been placed on group as opposed to individual-animal responses to a drug. This is surprising, since in the final analysis we are required to treat individuals and have little basis at present for predicting their response to drugs. Supplying the mean and standard deviation for a group response is important, but it is no substitute for studying and understanding the dynamics of responsiveness.

Statistically significant differences in the results of differing treatments frequently may reflect differences in response of the animals selected, not actual differences in the actions of the drugs. By the fortuitous selection, despite the use of randomization techniques, of animals which may be either very resistant to or very responsive to drugs (because of species characteristics, seasonal or sudden climatic changes, and so on), one may either overlook a useful compound or get an exaggerated notion of its activity. Expectations for response to a drug differ for animals with different base lines of activity. Thus, where it is possible to correlate base lines of activity for individual animals with their subsequent responses to drugs, one is in a greatly improved position for interpreting the data. Disregard of such differences may explain the discrepancies frequently noted between findings of different laboratories or investigators.

In studies of drug effects on locomotor and conditioned-avoidance behavior (*22, 23*), for example, in which correlations have been made for individual animals between their control levels of behavior and their subsequent responses to drugs, it has been possible to minimize or eliminate sex, intergroup, interanimal, and day-to-day differences as experimental variables and to greatly increase the precision of statistical analysis. Analysis of differences in response to drugs on an individual basis, in addition to providing information concerning the dynamics of responsiveness to drugs, can also be employed as an effective and more precise tool for studying and verifying the role of intrinsic factors (biochemical or physiological) which may control or influence the cellular response to drugs.

Analysis of Variables

One cannot but be impressed by the profound changes in response to drugs which can arise as a consequence of minor changes in an experimental procedure—changes in the intensity of a conditioned stimulus (*22*), the electrolyte content of a perfusate, the workload applied to muscle, and so on. The changes observed may be qualitative as well as quantitative, yet frequently there is insufficient attention given to this fact and failure to examine even some of our most time-honored procedures. For example, the variables in the widely used pharmacodynamic dog screen have never been adequately studied, even though, with this procedure, pharmacologists may at times draw conclusions from findings in only one or two animals. The same comment applies to measures of general locomotor activity, where the various devices used (jiggle-cages, revolving treadwheels, stationary photocell units, and so on) differently influence an animal's behavioral performance and responsiveness to drugs (*8*).

The study of experimental variables is time-consuming, but one can hardly expect to understand the meaning, reliability, or possible relevance of ac-

cumulated data without making a study of this kind. Such information enables us to establish optimal conditions for drug screening and evaluation and provides fundamental knowledge about the role of external factors in modifying individual or group responsiveness to drugs; and it may also provide insights concerning the mode of action of drugs. This is an area of study that has been much neglected.

International Standardization

The pharmacologist has been characterized in part as an "individual who never uses a procedure without modifying it," but it would be useful if he were to accept and develop the notion of establishing rigidly standardized procedures for at least some of his measures. For the purpose of better communication and understanding, particularly in the evaluation and comparison of new drugs, there is much to be gained from the use of internationally standardized procedures and equipment for widely used measures. In addition, this would seem to be a necessary first step toward a systematic study of the sources of interlaboratory variability. Although some collaborative studies have been undertaken to determine variability in certain specified procedures (*24*), definitive information on the sources of interlaboratory variability is still lacking. Unfortunately, such studies are unlikely to be undertaken unless a special group is organized with the responsibility of undertaking them. It is to be hoped that such a group will someday be established, and that pharmacologists throughout the world will cooperate with it. The results should greatly increase our understanding and also our ability to communicate.

Prediction from Animals to Man

In making predictions from animal studies to man it is assumed that many of the attributes of behavior found in man are also found in the higher animals, to a sufficient degree to cause man and the higher animals to be similarly affected by drugs. This working hypothesis has proved to be a useful one and appears to be as applicable to the psychologic as to the autonomic or neurologic effects of drugs, except when one is dealing with pathological conditions of unknown etiology which cannot be duplicated and studied in animals as such. Differences also are likely to arise because of the more complex behavior characteristic of man, particularly when we consider that the response to a drug is a result of a complex interaction of drug, tissue, personality, and environment.

Despite the handicaps and limitations implicit in animal studies, however, one can do a great deal more in predicting from them the dosage, clinical efficacy, side effects, and therapeutic ratio of drugs in man than is generally considered possible. To do this, however, will require a change in emphasis. It requires attention to *therapeutically relevant* measures of drug activity, and to the measurement of dose effects comparable to those sought or considered acceptable for man, for it is far easier to predict the probable clinical efficacy and side effects of a drug from a realistic base line than from the extreme base lines of activity on which most of our median effective dose (ED_{50}) values are based. To achieve this realistic base line, however, will require an increase in the sensitivity of many of our methods. Drug effects which can be observed only after the administration of doses producing unacceptable side effects or toxicity can hardly be considered to have therapeutic significance.

In addition, greater reliance should be placed on the unique faculties of the properly trained human observer to distinguish and quantify the desirable and the undesirable attributes of drugs and the relationships between them. Where procedures for observation and quantification are properly and systematically defined, one has available in the human an instrument capable of far greater quantitative discrimination than has been supposed, as well as a most effective and efficient laboratory tool for the simultaneous recording, collating, and integrating of many of the observations (in particular, the neuro- and psychopharmacological observations) essential to intelligent drug screening or evaluation. We must take this "instrument" seriously enough to develop it as a truly quantitative and reliable tool. The ability to predict, from animal studies, the effects of drugs on man will be greatly increased through attention to these factors.

References and Notes

1. C. Bernard, *An Introduction to the Study of Experimental Medicine* (Schuman, New York, 1927).
2. A. Smith, quoted by I. H. Freeman, in "Drug trade held a 'whipping boy,'" New York *Times* (10 Dec. 1959).
3. P. de Haen, *F-D-C Repts.* **22**, 16 (11 Jan. 1960).
4. A. Wikler, *The Relation of Psychiatry to Pharmacology* (Williams and Wilkins, Baltimore, 1957), pp. 268–270.
5. R. W. Houde, A. Wikler, S. Irwin, *J. Pharmacol. Exptl. Therap.* **103**, 243 (1951).
6. H. F. Fraser and H. Isbell, *ibid.* **105**, 458 (1952).
7. D. Mainland, *Clin. Pharmacol. Therap.* **1**, 411 (1960).
8. S. Irwin, *Can. Rev. Biol.* **20**, suppl., 239 (1961).
9. ———, in *Research Conference on Therapeutic Community*, H. C. B. Denber, Ed. (Thomas, Springfield, Ill., 1959), p. 29.
10. J. H. Burn and R. Hobbs, *Arch. intern. pharmacodynamie* **113**, 290 (1958); M. R. A. Chance, *J. Pharmacol. Exptl. Therap.* **89**, 289 (1947); J. A. Gunn and M. R. Gurd, *J. Physiol. (London)* **97**, 543 (1940); L. Lasagna and W. P. McCann, *Federation Proc.* **16**, 315 (1957).
11. B. B. Brown, in "Psychopharmacology: Problems in Evaluation," *Natl. Acad. Sci.–Natl. Research Council Publ. No. 583* (1959), p. 236.
12. V. Nowlis and H. Nowlis, *Ann. N.Y. Acad. Sci.* **65**, 345 (1956).
13. Standardized, multidimensional procedures have been developed at the Schering Corporation laboratories for quantifying and evaluating the behavioral, neurologic, autonomic, and toxic effects of drugs in both small and large animals. These procedures are based exclusively on gross observation and require no instrumentation; quantification is operationally defined. A brief description of the procedures was presented at a Gordon Research Conference on medicinal chemistry in 1959; the text of that report is available on request. A more complete description of the methods is in preparation. See also S. Irwin, *Arch. intern. pharmacodynamie* **118**, 358 (1959).
14. S. Norton and E. J. DeBeer, *Ann. N.Y. Acad. Sci.* **65**, 249 (1956).
15. S. Irwin, in *Psychopharmacology Frontiers*, N. S. Kline, Ed. (Little, Brown, Boston, 1957), p. 251.
16. E. W. Pelikan, *Pharmacologist* **1**, 73 (1959).
17. H. Riley and A. Spinks, *J. Pharm. and Pharmacol.* **10**, 657 (1958).
18. J. E. P. Toman, in "Psychopharmacology: Problems in Evaluation," *Natl. Acad. Sci.–Natl. Research Council Publ. No. 583* (1959), p. 175.
19. J. A. Bain, *ibid.*, p. 199.
20. E. K. Killam and K. F. Killam, *ibid.*, p. 159.
21. S. S. Kety, *Science* **132**, 1861 (1960).
22. S. Irwin, in *Dynamics of Psychiatric Drug Therapy*, G. J. Sarwer-Foner, Ed. (Thomas, Springfield, Ill., 1960), p. 5.
23. ———, M. Slabok, G. Thomas, *J. Pharmacol. Exptl. Therap.* **123**, 206 (1958).
24. C. I. Bliss, *J. Assoc. Offic. Agr. Chemists* **29**, 396 (1946); ———, *J. Am. Pharm. Assoc. Sci. Ed.* **38**, 560 (1949); L. C. Miller, *ibid.* **33**, 245 (1944).