
From the Discussion to the Fourth Symposium:

MEASUREMENT OF CHANGES IN HUMAN BEHAVIOUR UNDER THE EFFECTS OF PSYCHOTROPIC DRUGS

JOSEPH ZUBIN

Chief of Psychiatric Research (Biometrics), New York State Department of Mental Hygiene, and Professor of Psychology, Columbia University, New York, N.Y. (U.S.A.)

The coming of the drug age in psychiatry so soon after the heroic age of shock and lobotomy, has raised a challenge, not only for the psychiatrist, but also for the pharmacologist, psychologist and biometrician. Practice has already outstripped scientific knowledge, and the scientists working in this area are like reluctant dragons, pulled all unwillingly and unprepared into the fray. I shall let the psychiatrists and pharmacologists speak for themselves and limit my remarks to psychology and biometrics.

It is fortunate that Lehmann and Knight have focused their attention on the immediate effects of psychotropic drugs on the performance of normal subjects in psychological tests. Had they gone beyond that, it would have been impossible to deal with their topic. In contrast with previous workers, who have concentrated largely on the classification of drugs according to type of response, or of individuals according to sensitivity, the present authors have studied the tests themselves. They arrived at the following conclusions:

1. Tests may be classified as either placebo-resistant or placebo-prone.
2. Test reliability varies both diurnally and from day to day.
3. There are six overall parameters of test performance:
   a) tendency to increase performance (excitation)
   b) tendency to decrease performance (inhibition)
   c) drug sensitivity
   d) placebo-sensitivity (already mentioned)
   e) neuroticism and
   f) extraversion.
4. As a group, the scores tended to be normally distributed for neuroticism and extraversion; but increment (excitatory) and decrement (inhibitory) scores, as well as responsiveness to drug and placebo, showed bimodal distributions.

These conclusions are based on several studies ranging in number of subjects from 15 to 35. The authors point out the tentative character of their findings in view of the small samples.

In passing, the authors make certain suggestions about the relative merits of statistical treatment and clinical evaluation, and indicate that statistics can only be used to eliminate non-significant results, but not to establish significant results.

So much for the content of the paper. Now, for an analysis of its import.

It must be admitted at the start, that neither the psychologist nor the biometrician is happy about current tests and methods in pharmacological evaluations. We are really not ready, nor will we be ready for some time, with good tools. We are in somewhat the same position as the space scientists, who are also not prepared to tackle the major problems because they have not yet tooled up.

For this reason, Lehmann and Knight are to be congratulated on their attempt, even though it may not prove to be fully satisfactory.

Lehmann and Knight have attempted to provide a test battery, but we still

This study was supported in part by U.S.P.H.S. Grant No. 1541 from the National Institute of Mental Health.
don't have good tests to put into it. They are not to be blamed for this, for they point out the kind of needs that exist. Just as the clinician is usually ahead of the scientist in his hunches and intuitions, arising as they do from his daily battle with disease, so the clinical research man is ahead of his basic science colleague in pointing the direction in which research should go. Lehmann and Knight have gone beyond pointing the direction; they have actually begun to lay a road. But there are many obstacles in the way.

First, they as well as Pichot point out that the most important tools today are rating scales, rather than tests. This is correct, but subjective rating scales are the forerunners of measurement, even in physical science. Long before the thermometer was invented, clinicians could detect a fever by touching the patient's brow and still earlier, the ancient Egyptians rated subjective warmth on a scale of four steps ranging from the hottest day of summer to the coldest day of winter with two intermediate grades. I often wonder whether it wasn't a mere accident that warmth, height, weight, and time were objectified before we attained measures of depression and anxiety. Perhaps in one of the other life-bearing planets, the reverse order has been followed. But it is clear that ratings also can be made more and more objective and reliable. Apparently, the necessary element for transforming a subjective rating scale into an objective test is the discovery of an external criterion outside of ourselves, an expanding column of mercury instead of subjective warmth, a balance instead of subjective heaviness, rulers in place of arms and feet for length, intelligence tests in lieu of subjective comparison with the interviewer, etc.

Second, the authors contrast clinical evaluation with the objective evaluation of test results by pointing out that there are statistical criteria for eliminating tests leading to negative findings, but that clinical acumen is necessary to identify positive findings. Let me hasten to add that negative findings are just as liable to error as are positive findings. Negative findings lead to the well known error of the first type: erroneous rejection of a correct hypothesis; positive findings lead to the error of the second type: erroneous acceptance of a false hypothesis. But actually there is no opposition between clinical and experimental findings. No one would want to stop the clinician from making a guess about the value of a given statistical finding. However, that hunch should not be taken as final, but should be subjected in turn to an experimental and statistical verification. That is in fact why experiment was started in the first place to test the tenability of a clinical hunch. Thus, the clinical vs. actuarial conflict is a pseudo-conflict to be resolved by a series of sequential experiments. It might be well to remember that both the clinician and the experimentalist deal with the subjective, but the clinician sometimes forgets the subjective nature of his material.

Third, identification of placebo-prone vs. placebo-resistant tests is indeed a worthwhile undertaking. But in this case the differences can as readily be ascribed to practice-prone vs. practice-resistant tests. Improvement with repetition is a characteristic of nearly all test performance. Whether it is due to learning, adjustment to testing conditions, elimination of feeling of strangeness, or what, is difficult to tell. It is interesting to note that the speed tests are the ones that the authors regard as placebo-prone, while power tests which demand accuracy of performance, are regarded as placebo-resistant. These are also respectively practice-prone and practice-resistant. The number of subjects who exceed the average increase under placebo is merely a
measure of skewness and not a good measure of the relative influence of the placebo. How to get a relative measure of placebo- or practice-proneness remains an important question that will have to be resolved in other ways.

Fourth, the reliability of the tests raises some disconcerting problems. Test-retest reliability is only one aspect of the problem; the authors have not touched on the problem of internal consistency.

It is doubtful whether the contrast between the correlations obtained the same day and those obtained on different days, as shown in the tables, is significant. Nevertheless, the well known diurnal variation in body temperature may well be reflected in the correlations between test and retest given at different hours, especially if the individuals in the group are not homogeneous with regard to the pattern of diurnal variation.

Fifth, the attempt to determine whether a given test is sensitive to inhibitory or excitatory stimulation is interesting, but far from convincing. For one thing, the particular test under scrutiny (for which the correlation is being determined) should be excluded from the change score, since comparison of a whole score with a part of itself gives rise to spurious correlation. A better method might have been to contrast the gains and losses for secobarbital (the inhibitor) with dextroamphetamine (the stimulator). This would have given a much clearer indication of whether or not a given test is sensitive to either factor. In a study which LANDIS and I conducted on antihistamines, we used two anchoring drugs, phenobarbital and dextroamphetamine, to determine in which direction the unknown antihistamine tended. By setting up standard anchors with standardized inhibitors and stimulants, a continuum could be established on which a new drug could be placed.

Finally, the decrement in test performance used for measuring inhibition, the increment used for measuring stimulation, and the sum of increments and decrements, regardless of sign, used for measuring sensitivity to drug or to placebo, all represent highly correlated factors, since they all involve the same decrements and increments in performance in some way. Correlations between such measures must perforce exist since they involve identical elements. More independent methods must be found for measuring the factors in question.

But there is still another problem which faces us. The tests described in this paper were selected because they seemed to promise clinical usefulness. This is hardly a scientific way of selecting tests. Some schema is needed for classifying human behavior according to the likelihood of drug influence. Some time ago, faced with a similar need for classifying human behavior with respect to prognostic indicators in schizophrenia, we devised a classification schema in which each type of response—physiological, sensory, perceptual, psychomotor and conceptual—was arranged along one dimension of a two-way table, whose other dimension comprised the methods of eliciting such responses. The stimulus methods ranged from the idling state (i.e. no stimulus) through energy and signal stimuli to symbols.

This schema recognizes that the performance of a subject under a given stimulus gives rise to responses on all behavioral levels from the physiological to conceptual. The experimenter usually selects only one response level for investigation. Thus, presenting certain symbols to elicit conceptual behavior may also evoke physiological, sensory, perceptual and psychomotor responses; but in the usual experiment, these are neglected.
Drugs may be regarded as loads on the system, while performance in the idling state provides a base line for comparison. It is clear, however, that under the load of a drug, performance at other levels that are not under scrutiny during the idling state may become important. For example, the stimulus may be a light signal for which the response to be observed is psychomotor. Nevertheless, physiological, sensory, perceptual and conceptual activity set in motion under the influence of the drug may profoundly modify the psychomotor responses. Thus, fear engendered by the drug, or perhaps eliminated by it, may influence the psychomotor response under scrutiny. This, in fact, may be the basis for the placebo effect which may operate in either direction to yield an increment or a decrement.

While this schema has some arbitrary features and will probably be found lacking in some respects, it does provide a framework, a sort of Mendeleef table, from which to sample representative functions.

The problem of the placebo-response is one of the most difficult to resolve. We have already seen how Jellinek and Beecher and Lasagna have gone about detecting the placebo reaction.

From the point of view of this table, a drug or placebo is best viewed not as a stimulus, but rather as a load on the organism which modifies the response of the individual to any stimulus and which may bring about changes at all the levels of behavior. Individuals whose physiological, sensory, perceptual, psychomotor and conceptual responses are not affected by placebos apparently compensate for the load.

Some confirmation of this point of view is provided by Joyce. He found that placebo reactors could be differentiated from non-placebo reactors by means of three tests administered before the administration of the placebo: (1) selected items from a personality inventory (Bernreuter); (2) awareness of ongoing autonomic activity, e.g., digestion, respiration, heart beat, as revealed by a questionnaire (Mandler, Mandler and Uviller); and (3) the actual response to the application of a placebo. The first two of these tests are of conceptual attitudes which may indicate greater sensitivity to the expected effects of drugs, rather than the actual effect of the drug. It is interesting to note that the objective psychological performance tests used in this study did not differentiate the placebo reactors from the non-placebo reactors.

Why individuals who are consciously aware of their autonomic behavior are more sensitive to placebos is an interesting question to consider. It is possible that overt awareness of autonomic activity leads to greater and easier conditioning of these internal events. If so, the past experience of the person which has conditioned his responses to ingested chemicals may be the basis for placebo reaction. In fact, the role of conditioning in producing reactions to placebos as well as to drugs needs to be investigated.

As regards the statistical treatment of data, some voices have been raised recently against the use of the subject as his own control. The development of statistics which are suitable for treating each individual as a separate universe is the crying need of the hour, since each individual has his own unique biochemical structure, personality characteristics and sensitivities to drug or placebo. This is not to say that there are no regularities from person to person. But we must know the personal parameters before these regularities can emerge. Once these personal parameters are known, like-structured individuals can be sorted into groups.

In summary, we owe the authors of this paper a vote of thanks for opening the
door to problems in urgent need of solution. The authors have presented their solutions as tentative. I have gone on to point out some needed revisions. But the eventual solutions will have to come slowly. We must follow along the path of Galileo who, in the words of Dingle: "... left the world of angels and spirits until the time should come when it could be explored, and contented himself with such principles as he could extract with confidence from experience, though the resolution committed him to such trivialities as the timing of balls rolling down grooves. It is that self-control—the voluntary restriction to the task of extending knowledge outwards from the observed to the unobserved instead of imposing imagined universal principles inwards on the world of observation—that is, the essential hallmark of the man of science, distinguishing him most fundamentally from the non-scientific philosopher."

Step by step, piece by piece, we must build our troughs and roll our balls down their grooves. Test batteries, though necessary, are like shotguns. They achieve their effects because of wide coverage. Further progress requires careful study of each level of behavior through the investigation of individual functions.

Unfortunately, Dr. Pick's paper did not reach me until the day of the meetings and I can give only the results of a hurried survey of its contents. It covers the literature of drug evaluation with particular reference to the following considerations: (1) the situations in which drug evaluation is called for; (2) the instruments used; and (3) the methods of determining the direct effects of the drug.

Regarding the situations in which drugs are to be evaluated, he contrasts work with normals and abnormals, pointing out that some proportion of normals usually possess psychopathological characteristics, requiring careful screening. How this screening is to be done objectively and validly is still a moot question. The administering of drugs to mental patients raises the ethical issue of whether the drug in question is as efficacious as other methods already established in practice. The issue is a thorny one; but the outcome of research can never be guaranteed, and to refrain from doing something new because of such hazards would eliminate all progress.

The instruments to be used in drug administration are divided into two major divisions: clinical and psychometric. The clinician makes use of global and specific methods and rating scales. It is important to recognize that often his aim is not identical with the aim of the scientist. When global clinical evaluation is used for the evaluation of a drug, only gross clinical changes can be observed, while the more subtle subclinical changes of interest to the scientist go unnoticed. In an attempt to reduce the burden of global clinical evaluation into its underlying components, certain focal symptoms are often selected for observation; and again subclinical changes are likely to be overlooked. In other words, objective refined measuring instruments are needed to supplement the gross clinical observation to guarantee that these more subtle changes do not escape notice. While these subtle changes may not be of immediate clinical significance, they may point the way to improving the drug's efficacy by altering its chemical structure, or its dosage. Thus, rating scales may sometimes reveal changes which the clinician fails to notice because of the subclinical nature of these changes.

It is generally agreed that at the present time, rating scales are far superior to psychometric tests for picking up behavioral changes. Nevertheless, a sampling of behavioral changes in line with the framework indicated on page 335 may reveal changes that now go unnoticed.
The rating scales, to be effective, must depend upon direct observation of overt behavior and eschew dynamic interpretations. The latter better be left to the global clinical evaluation. Such behavioral scales should be used by that member of the hospital staff who observes directly the behavior in question. Thus, questions about sleep should be answered by the night nurse or attendant, questions about work performance, by the occupational therapist or work supervisor. Rating scales are now available for rating the behavioral characteristics of the patient in the ward, in the work-shop, during testing with the psychologist, during interview with the psychiatrist, during interview with the social worker and during interaction with the family. These scales give an objective picture of patients behavior in all of these situations and their patterning is of extreme usefulness to understanding the significance of the changes observed.

Only one critical comment arises regarding the use of rating scales. Too often, a rating scale is translated from the language of one country to that of another and used as if the cultural differences between the two countries could be dismissed. The great temptation to do this should be resisted and only after a new standardization has been carried out in the new country should the scale be applied. Unfortunately this is expensive, and I often wonder whether it might not be better for each country to develop its own rating scale in line with its own tradition and cultural expectancies. This would make it impossible to make cross-cultural comparisons, but cross-cultural comparisons with translated instruments may turn out to be spurious. An international board of experts might profitably be convened to discuss this matter.

One issue which Pichot raises requires further consideration. Is it necessary to have a theory of personality before one proceeds to measure changes in human behavior? No one will deny the usefulness of personality theory and testable hypotheses derived from it as a basis for experimentation. The only question that is before us is, are we ready for such a personality theory? I believe that we are not yet ready and any attempt at foisting an a priori personality theory on our observations is bound to redound to the discredit of our efforts. We should be more modest and satisfy ourselves with limited specific hypotheses regarding the effect of drugs, and sample the various levels of behavior likely to show changes under the effect of drugs. When more is known about such changes, perhaps a personality theory might be built on the limited specific hypotheses that prove to be tenable. In the last analysis, however, each scientist follows his own preferences in this regard and perhaps the interaction between the global theorists and the specific empiricists is necessary for progress.

In conclusion, both contributors have rendered the field of psychopharmacology a great service by laying bare the series of problems that face the experimenter in this field. These efforts will bear fruit in so far as they attract enough attention to bring about the solution of the problems they have raised.

REFERENCES


Printed in The Netherlands