Meehl, P. E. (1967). What Can the Clinician Do Well? In Douglas N. Jackson and Samuel Messick (Eds.) Problems in human assessment (pp. 594-599). New York: McGraw-Hill

What Can the Clinician Do Well?

PAUL E. MEEHL

In the preface to the book *Clinical versus Statistical Prediction* (Meehl, 1954) which was, in part, responsible for this symposium, I wrote that students reacted to my lectures on prediction as to a projective technique. Many psychologists have responded to the book in the same way. I am therefore going to take this opportunity to repeat, with refinement and clarification, the statement of my essential position, reserving for another time at this Convention the presentation of empirical material from my current research.*

First of all, I am puzzled by the extent to which both statisticians and clinicians perceive the book as an attack upon the clinician. On the contrary, my position was, and is, that the clinician performs certain unique, important, and unduplicable functions, in some of which he has literally no competition. I think the book states this very clearly. (I hope it's true, since I occupy almost one-third of my time and earn a sizeable part of my income in clinical work!) But in current practice, clinicians spend a good deal of time and energy performing functions at which there is neither theoretical nor empirical reason for supposing them to be efficient. My position is not, therefore, one of being "for" or "against" the clinician, or proposing to eliminate him. I cannot understand, for example, how my friend Bill Hunt could possibly read me as viewing "… the exercise of clinical judgment as a necessary evil," as he states in the Bass and Berg volume, rather than, as for him, "… a fascinating phenomenon with a genuine predictive potential." Two full chapters of my book, and portions of two others, were devoted to analyzing (and defending!) the clinician's non-formalized judging and hypothesizing behavior, and I should have thought that my own fascination with the phenomenon was quite apparent.

However, I did want to influence clinical practice toward a more optimal utilization of skilled time, by removing the clinical judge from loci in the decision process where he functions ineffectively, thereby both (a) improving predictive accuracy and (b) freeing the clinician's time for other activities, whether cognitive or manipulative, in which he is efficient or unique.

Some feel that it was a disservice to formulate the problem in terms of opposition or competition, as clinical "versus" statistical prediction. Reading their discussions as sympathetically as my own bias permits, I remain persuaded that a pragmatically meaningful decision problem, involving a comparison between two distinguishable procedures, does exist. Discussions which have appeared during the last five years do necessitate some reformulations but, although I may be overly identified with my original position, I cannot regard them as fundamental. Given a set of data on a patient, and given the pragmatic necessity to make a certain decision, one may *either* combine these data, or selected portions thereof,

*This article is published with the permission of the author. It was originally read at the symposium "Clinical Skills Revisited," September 4, 1959, American Psychological Association Convention, Cincinnati, Ohio.

in a formalized, mechanical, clerical fashion (of which the regression equation is only one example, perhaps not the most powerful); *or* he may invite a clinician, or a staff conference, to think and talk about these data and come to a decision. Now no one contends that the individual decisions resulting from these two methods of combining data will always coincide. This would be theoretically preposterous; and there is, of course, a massive body of experience to the contrary. In those numerous instances in which they *fail* to coincide, one must act in some way. If he acts in accordance with the decision provided by the clerical procedure, he has countermanded the judgment of the clinician or staff conference. If he acts in the other way, he has countermanded the statistical formula. *I have yet to see any cogent, or even plausible, criticism of this fundamental point, which was made clear by Sarbin 15 years ago.* It is thoroughly misleading to speak of Sarbin or Meehl as fomenting a controversy, or as having set two procedures in "needless opposition" to each other. They *are in* daily opposition, manifest or covert; their opposition is an immediate logical consequence of a simple, undisputed fact: namely, that the human judge and the statistical clerk correlate less than 1.00 when required to make diagnoses, prognoses, or decisions from a given body of information. Nothing is gained by adopting a hysteroid, "sweetness-and-light" attitude, akin to Mr. Dooley's definition of democracy as a situation in which "everybody is equally better."

There has been an overemphasis upon the chapter in which I surveyed the then available empirical studies, some readers reacting with glee to the box score, and others stressing the fact that these studies are not ideal, as I also emphasized in summarizing them. Personally, I consider that chapter to be almost the least important part of the book. On the other hand, no detailed, rigorous analyses or criticisms of the *theoretical* and *methodological* considerations raised in the book have appeared. I am at a loss to know whether this is because everybody agrees with me, or because these considerations are mistakenly thought to be irrelevant. I am convinced that the *formal* arguments on the actuarial side are very powerful, and they ought not to be thus airily dismissed or by-passed.

Well, what can the clinician do well? However well or badly he does certain things, he alone can do them, and therefore it is administratively justifiable to occupy his time with them. He can, for instance, observe and interview the patient, functions which are not eliminable by any kind of statistics. He can *be* a person himself in relation to the patient, with all that this means for the helping process. He can construct hypotheses and carry out research to test them. Every hour saved out of those innumerable and interminable staff conferences and team meetings in which some clinicians seem to delight (you know, there are clinics where the average weekly hours of inter-staff contacts exceed those of staff-patient contacts!) can be devoted to seeing patients and doing research.

Among decisions which can, in principle, be arrived at either by a formalized or a judgmental method, I would now state the generalized clinical-statistical issue something like this: "Given a population of patients, with variable information on each; and a population of judges, with variable information on each; and a decision task imposed upon us pragmatically; *then,* at *which* points in the total decision-making process should we use *which judges;* and at which *other* points should a non-judgmental ('formal,' tabular, graphical, statistical, clerical, 'mechanical') operation be employed?"

It astounds me that, in spite of my having very carefully distinguished between *type of data* (i.e., psychometric or non-psychometric) and *method of combining data* (i.e., judgmental or formalized), numerous writers have continued to perpetuate the old confusion between these two; several have even quoted me as maintaining that tests are better predictors than non-test data! This is a remarkable projective distortion, especially since I am extremely skeptical myself as to the predictive power of the available tests in the personality field. I have held for some years that life-history and "mental status" variables are probably superior to existing tests, a superiority which I expect to become clear as actuarial methods of combining these non-psychometric data are more widely utilized.

In answering the general question for a given prediction problem, we must include the utilities of the several outcomes of right and wrong decisions, the cost (monetary and otherwise) of the alternative decision methods, and the distribution of hit-frequencies. "Equal hits" means "equal predictive success" only if hits are equally important whichever kind they are; and "equal predictive success," in turn, means "equal efficiency" only if one predictive method costs no more than the other.

The formulation is generalizable over all of the clinician's cognitive activity, whether predictive, postdictive, or diagnostic. Some have argued that the clinician doesn't "merely predict," but tries to influence the course of events, so that the problem posed is of little practical importance. This argument is philosophically naive. Selecting a certain line of action in order to influence the course of events is itself justified by implicative statements of the form, "If procedure X is carried out, the patient will respond in manner Y." And this, of course, is a prediction whether realized or counter-factual. A related error is made by those who have suggested that the clinician doesn't predict directly, but decides, upon the data he has, what additional data he needs before predicting. Not to act at a given moment in time, but to collect additional data of a specified kind is, of course, itself a decision; and is, like other decisions, rational or irrational, depending upon the probabilities and utilities involved.

In reporting a recent empirical study (Meehl, 1959) I listed 6 factors or circumstances theoretically favoring clinical prediction. Although that list was not presented as exhaustive, I have not yet come across any examples, either factual or armchair, falling outside these six rubrics:

1. *Open-endedness:* It often happens that the predictive task is not presented in the form of a prespecified criterion dimension or exhaustive set of categories, but rather as an open-ended question where the very content of the prediction has to be produced by the predictor.
2. *Unanalyzed stimulus-equivalences:* Sometimes the scanning and classifying of the data, including the recognition of a certain fact or pattern as relevant, cannot proceed by explicit rules because the operative "rules" are laws of our mental life as yet unknown or incompletely known. Perceptual gestalts, psychological similarities in physically dissimilar events, analogical and primary-process thinking, and similar inexplicit psychic processes are available to the predictor because he, being human, exemplifies laws which he may not be able to report because research has not yet elucidated them.

3. *Empty cells:* From time to time the prediction situation presents special cases in which a factor or configuration is highly relevant but has not occurred even in the course of very extended actuarial experience. In such cases the human judge must spontaneously notice the special circumstance and assign to it an estimated weight. In extreme instances such rare factors must be treated as "stop" items, being allowed to countermand an otherwise strong prediction reached by the formal (mathematical) procedure.
4. *Theory-mediation:* When a prediction can be made by the use of hypothetical mental constructs whose laws (usually very imperfectly known) are in such general form as to permit a variety of structural-dynamic arrangements *in concreto*, the predictive process is not straightforward because hypothesis-building is a creative, synthetic act for which automatic rules cannot be written. The $\text{fact}_1 \rightarrow \text{fact}_2$ sequence can always (in principle) be reduced to an actuarial generalization holding between members of the large (but finite) set of combinations and hence can be treated formally; whereas the $\text{fact}_1 \rightarrow \text{construct} \rightarrow \text{fact}_2$ sequence can not always be thus formalized. The extreme case of this situation is the rare one in which the clinician actually invents new nomothetic *constructs* (as distinguished from thinking up new concrete exemplifications of familiar ones) in formulating a particular case. Freud's early analyses exemplify this case.
5. *Insufficient time:* In some predictive situations (e.g., interpretive psychotherapy) the pragmatic context requires that the prediction, to be of any use, must be reached in a very short time, even a matter of seconds, after the relevant data appear. A therapist cannot put his patient in cold storage while he, the therapist, runs off a P-technique factor analysis on a 28-variable correlation matrix derived from the patient's verbal productions during the preceding 30 minutes. Even if every office of the ten thousand skilled therapists in the U.S.A. were somehow provided with a high-calibre electronic computer, the time required for coding and feeding would make this science fiction fantasy an inadequate solution.
6. *Highly configurated functions:* Suppose that a configural relationship exists between a set of predictor variables and a criterion, but that the function is not derivable on rational grounds. We have to approximate this unknown optimal formula by empirical methods. Multivariate tests such as the Strong, MMPI, and Rorschach provide familiar instances of the problem. Clinicians skilled in the use of these devices find it helpful to have the several scores expressed graphically as a psychogram or profile, and this practice is not merely a matter of convenience in reading. Typically the clinician reports that his inferences from the profile are based partly upon discriminations he has learned to make among the various "patterns" which arise in an extended clinical experience. Usually these patterns are grouped into categories or types, but the clinician recognizes the existence of numerous intermediate forms so that the underlying function is presumably continuous. What seems to be happening is that an unknown configurated mathematical function is being approximately expressed via the graphical mode, utilizing the fact that differences and

similarities of visual gestalten can be perceived without the percipient's knowing the underlying formula.

Each of these six presents its own special problems for research, and in most actual clinical judgments more than one is likely to be operative. Like Dr. Hoffman, I have chosen to investigate the sixth factor empirically, although I would readily agree that it is theoretically the least interesting. It has, however, the advantage of being somewhat easier to subject to quantitative analysis, and sizeable samples of such judgments are fairly easy to obtain.

Admittedly, a mere tally of the "box score" based upon heterogeneous studies comparing the efficiency of formal and judgmental methods of combining data is not as helpful, either for practical purposes or in giving us greater theoretical insight into the clinician's cognitive activity, as will be systematic studies of these six components as they appear at different stages of the total decision-making process. Where, in this chain of gathering facts and making inferences, is the skilled human judge indispensable? Where is he dispensable, but only with a loss in predictive accuracy? Where are his cerebrations inferior in their outcomes to the application of a formalized procedure, such as an actuarial table or mathematical equation? Many studies, in the several domains of predictive activity, will be needed to answer these questions.

However, since large amounts of time are being spent today making decisions by impressionistic, judgmental, and conversational methods (such as the staff conference), it is worthwhile to attempt a rough generalization as to the relative power of the two methods over diverse predictive domains. I cannot agree with those who consider that such a box score is either meaningless or unimportant. By the latest count, there are 35 studies which permit a comparison between the human judge and a clerical, formalized procedure for combining information. Many of these studies are not based upon clinical data of a high order, either in quantity or quality. However, I must emphasize that *many of the studies do involve amounts and quality of data quite comparable to what are routinely available in most clinical and counseling situations and which are being applied daily by clinicians in making their judgments.* The shortage of skilled professional personnel, which is certain to be with us (and in fact to get worse) during the lifetime of everyone in this room, makes it thoroughly unrealistic to argue that no significant comparison can be made unless it involves the kind of workup that a wealthy patient in a plush mental hospital receives at fancy prices. I would further point out that it is a quite unjustified assumption, commonly made by critics of the box score, that "naturally," if the quality and quantity of the clinical data, and the professional competence of the clinicians, were deliberately picked as being of a very high order, the clinician *would* show up markedly better than a souped-up, configural statistical prediction system *utilizing the same top-quality data.* This may or may not be true; some of us are still patiently watching the journals for the evidence.

Of the 35 studies known to me, 12 deal with predicting outcome in some kind of training or schooling; 8 with recidivism, delinquency, or parole violation; 3 with improvement of psychotics; 3 with psychiatric diagnosis, i.e., the attachment of a nosological label; 3 with the outcome of outpatient psychotherapy of neurotics; and 5 with personality description not covered by any of the preceding such

as Q-sort characterization of a patient, aggression as inferred from the Rorschach, and the like. One study compares the two methods in organic medicine. If we define equal efficiency without regard for time and economics (a definition strongly biased against the statistical technique), we find 23 of the 35 studies showing a difference in favor of the statistical method, 12 studies showing approximate equality, and no study favoring the judgmental method. Of course from the social and economic viewpoint, this really means 35 studies on the actuarial side. The overall picture has, therefore, not changed since 1954, except that the proportion of "equal" outcomes has somewhat decreased.

I think that it is time for those who resist drawing any generalization from the published research, by fantasying what *would* happen if studies of a different sort *were* conducted, to do them. I claim that this crude, pragmatic box score *is* important, and that those who deny its importance do so because they just don't like the way it comes out. There are very few issues in clinical, personality, or social psychology (or, for that matter, even in such fields as animal learning) in which the research trends are as uniform as this one. Amazingly, this strong trend seems to exert almost no influence upon clinical practice, even, you may be surprised to learn, in Minnesota!

In the single study of medical diagnosis, it was found that a linear discriminant function combining the results of two biochemical tests did as well in differentiating types of jaundice as did internists who, in addition to these two tests, had available a large mass of other information, and averaged between 3 and 4 hours going over each patient's material! Some psycho-clinicians oppose the actuarial method on the ground that physicians have been practicing medicine for centuries without it. This argument completely mystifies me, since, with the exception of this one study, no comparison of the two methods in organic medicine has ever been made. The frequency of erroneous diagnoses in medicine is well known, and it is hard to imagine why anyone familiar with organic medicine would give such an argument any weight.

There are physicians who have begun to apply statistical techniques, the mathematics of decision theory, and electronic computers to medical diagnosis. Those psychologists who use the analogy with medicine counter-actuarially would be well advised to wait until we find out what happens there. It would be ironic indeed (but not in the least surprising to one acquainted with the sociology of our profession) if physicians in non-psychiatric medicine should learn the actuarial lesson from biometricians and engineers, while the psychiatrist continues to muddle through with inefficient combinations of unreliable judgments because he has not been properly instructed by his colleagues in clinical psychology, who might have been expected to take the lead in this development.

I understand (anecdotally) that there are two other domains, unrelated to either personality assessment or the healing arts, in which actuarial methods of data-combination seem to do at least as good a job as the traditional impressionistic methods: namely, meteorology, and the forecasting of security prices. From my limited experience I have the impression that in these fields also, there is a strong emotional resistance to substituting formalized techniques for human judgment. Personally, I look upon the "formal-versus-judgmental" issue as one of great generality, not confined to the clinical context. I do not see why clinical psychologists should persist in using inefficient means of combining data just because

investment brokers, physicians, and weathermen do so. Meanwhile, I urge those who find the box score "35:0" distasteful to publish empirical studies filling in the score board with numbers more to their liking.

References

Meehl, P. E. (1954). *Clinicial versus statistical prediction.* Minneapolis: University of Minnesota Press. [Reprinted with new Preface, 1996, Northvale, NJ: Jason Aronson. Reprinted 2013 by Echo Point Books.]

Meehl, P. E. (1959). A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles. *J. counsel. Psychol.*, *6*, 102-109.

pdf by LJY, March 2016