PSYCHOLOGICAL PROGNOSIS OF OUTCOME IN THE MENTAL DISORDERS

JOSEPH ZUBIN AND CHARLES WINDLE
Psychiatric Institute, Columbia University

When mental disorders are regarded as continuing processes running a regular course with definite characteristics and not as haphazard vicissitudes of life, it becomes possible to study the initial, middle, and ultimate stages of these processes and hopefully to obtain indicators in each stage of the probable subsequent course of the disorder. This is the essence of the prognostic problem. The personality of the patient, his genetic and environmental background, his family relationships, vocational and social adjustment, in fact all his assets and liabilities at any given period, may become important indicators of the probable course of his illness from then on.

The wise clinician has always been aware that the characteristics of the patient at the time he comes for treatment have an important relationship to his chances for eventual improvement, provided the stage of the illness can be determined. One able diagnostician, Kraepelin, has even proposed for one illness, dementia praecox, that diagnosis be based on the probable outcome of the disease. Present-day clinicians still follow this pattern to some extent; they examine disease processes with an eye to their eventual outcome and diagnose accordingly. But the observations on which these prognoses are made are usually impressionistic. Whether the psychologist can do any better with tests than the clinician with his intuition is perhaps debatable, but there is such a definite demand for improving prognostic criteria that tests are well worth trying. The demand for improvement arises in the wake of the general finding that modern specific therapies of both the psychotherapeutic and somatotherapeutic variety seem to have an advantage over nonspecific therapies, such as "total push," only for the period immediately following the therapy. In long-range follow-up studies, no conclusive evidence is available for any advantage in favor of specific therapies (84). There are two possible explanations for the failure of the specific therapies to stand up under long-range scrutiny: (a) the specific therapies help only those patients who would improve anyhow, and (b) each of the specific therapies is effective only for a specific type of patient. Perhaps, if prognostic tests were available to predict the different outcomes to be expected for a given patient under each of the specific therapies, better results than those afforded by nonspecific therapies would follow. In either case, prognostic information seems highly desirable. Consequently, psychological tests have been employed with increasing frequency in order to provide more standard indications of probable outcome, and this has made possible confirmatory studies and statistical evaluation of the reliability and validity of the measures.

Despite the great increase in interest in prognosis today, most studies have preferred to follow the pattern laid down by the earlier diagnostic studies rather than blaze new trails which would lead more directly to the goals of prognosis. There are some problems common to both diagnostic and prognostic studies, such as, (a) procedures for establishing validity of instruments; (b) classification of patients into homogeneous categories such as age, sex, etc.; (c) proper statistical treatment of the data. Among the problems peculiar to prognosis are: (a) proper design of prognostic study; (b) specification of therapeutic agent or type of treatment to which patient is to be exposed; (c) duration of illness at time when prognosis is made; (d) stage of illness at which prognosis is made; (e) duration of follow-up; and (f) criteria for evaluating outcome. We shall deal with each of these difficulties in turn.

The psychological tests used to evaluate functioning of patients cannot now be validated directly but must depend for their ultimate
validation on interviews and direct observations of patient behavior. In contrast with physiology, which also evaluates functioning, psychology has no anatomy to which the observed behavior can be related. The so-called structural elements portrayed in Freud's anatomy of the mind, Lewin's "differentiated regions of the life space," Rank's will, McDougall's sentiments, Stern's self, Spearman's and Taurstone's factors, and the personality structure conjured by Rorschach and projective technique workers, are little more than an assemblage of abstractions derived from descriptions of behavior, rather than independent constructs amenable to direct measurement. Perhaps the neurophysiologists will some day perfect models of brain function which may serve as a physiological substratum for perception, thinking, and action, but such models are still in the future. Until that time, the only possible way to validate psychological tests is through direct observation of behavior and interviews.

Prognostic studies, unless well designed and appropriately implemented, are of little value. Too often the prognostic aspects of a study are introduced as an afterthought when the major aim of the study has failed to materialize. In order for a prognostic study to be scientifically acceptable it must be planned for, and must satisfy certain elementary criteria in its presentation of the data. Of primary importance, of course, is the specification of the particular pathological condition under investigation, including both the specific disease and the condition of the patients with regard to age, sex, attitude toward illness, etc. Also important are the form of therapy and the time relations in the prognosis. It is well known that the stage of development of the illness is relevant to prognosis. The outcome immediately after therapy is often at variance with the outcome after several years, and for this reason the period of follow-up must be specified. One might think that these elementary requirements of scientific reporting are so self-evident that they hardly need mentioning. Unfortunately, many of the studies in the literature fail to provide these crucial bits of information. The use of suitable statistics for evaluating outcome is, of course, essential in prognostic studies. Nevertheless, some studies in this field suffer from the inadequate use of statistics, many being content with an impressionistic comparison of central tendencies. Furthermore, most of the prognostic studies today are based on hindsight rather than foresight. It is to be hoped that as more experience is gained, a few hardy souls will attempt truly predictive studies.

Another difficulty that hampers prognostic studies is the absence of uniformly acceptable criteria of improvement against which prognostic tests could be validated. Such terms as "cured" and "recovered," appropriate as they may be in acute physical diseases like appendicitis, are not at all applicable to chronic illnesses like schizophrenia. The goal of therapy, which is fairly well defined in the acute physical diseases—to restore the patient to his premorbid status—must be readjusted to a lower level of aspiration in the case of chronic illnesses (82). As a realistic basis for the evaluation of outcome of therapy one must still resort to such crude measures as being in or out of the hospital.

In view of the above difficulties, the only method of ascertaining whether any relationships exist between test performance and subsequent outcome was to search for possible consistencies in the various studies despite their shortcomings. Previous reviews of the literature on this topic (16, 18, 64, 66) have been of limited scope and have not attempted a thoroughgoing resolution of the conflicting claims made for the various tests. Consequently, a more complete review (78) was undertaken by one of the present authors to determine the validity of the prognostic claims for each of the psychological tests used.

It was found that most of the reports of the prognostic efficacy of psychological tests were concerned with three techniques: the Rorschach, the Minnesota Multiphasic Personality Inventory (MMPI), and a variety of intelligence tests. Each of these techniques had been claimed to be effective in predicting outcome, but in each case several investigators failed to confirm the claims. In addition, a wide variety of specific indices had been claimed to be predictive in a given test, and there was considerable diversity in the prognostic value or direction of any particular sign or index. This discordance can best be illustrated by presenting summaries of the findings employing these specific psychological tests.

The review of the literature disclosed 22 studies in which the Rorschach technique was
investigated as a prognostic agent in the psychoses.* All but seven (18, 29, 60, 61, 63, 66, 76) of these studies reported the Rorschach to have predictive efficacy. Furthermore, most of the studies claimed that in general the more capable patients or those who were most nearly normal were more likely to have favorable outcomes. In spite of this agreement in the over-all interpretation of the Rorschach protocols, the empirical signs proposed for prediction exhibited little consistency.

For example, one outstanding exponent of the prognostic use of the Rorschach, who has contributed eight papers to this field (38, 50, 54, 55, 56, 57, 58, 59), has proposed in each of seven different experiments a different set of signs or indices, often varying considerably from the signs reported in previous investigations. It appears that not one of the indices which was found to have prognostic power retrospectively, maintained this power in a new sample. The same inconsistency that occurs in the empirical findings is also present at the level of interpretation. Thus, running parallel with the seven different groups of prognostic indices are three slightly different interpretations of the personality patterns which are claimed to differentiate outcome groups (55, 57, 58). One can sympathize with the strivings to improve the test results, but one must also realize that the continued change of indicators may reflect the inappropriateness of the selected signs. Furthermore, the multiplicity of possible indices afforded by the Rorschach guarantees that some pseudosignificant differences will always present themselves.

Another example of the difficulties of the Rorschach technique is provided in a follow-up study by Sloan (73) of mental defectives discharged on wage placement. Seven criteria which had been suggested by another well-known Rorschach expert as prognostic for good adjustment were computed for contrasted outcome groups. Although the Rorschach criteria for good adjustment adequately differentiated the outcome groups, the predicted outcome was the very opposite of that which actually occurred. Those mental defectives who were successful in staying out on wage placement had a reliably smaller number of agreements with the criteria for good adjust-

* These 22 studies included 18, 20, 22, 29, 30, 38, 40, 47, 49, 50, 54, 55, 56, 57, 59, 60, 61, 63, 66, 67, 75, 76.

ment than did those who were returned as failures.

Additional difficulties in accepting the claims for the prognostic value of the Rorschach arise from the inadequate evaluation of the results of experiments in this field. Only six of the studies making prognostic claims with psychotics (40, 47, 54, 56, 59, 67) have utilized statistical methods to test significance, and two of these (47, 54) omitted the use of Yates's correction in the chi-square test with small samples. The best clinical verdict one can give regarding the prognostic claims of the Rorschach technique is: not proven.

It might be expected that studies of the prognostic value of the MMPI would show more consistency than those of the Rorschach. The scoring of this test is entirely objective, there are fewer standard indices that might appear to have prognostic value so that pure chance results play a smaller role, and less retrospective manipulation of the indices in the form of fractions, ratios, and weighted scores is possible. Consequently, it is to be expected that most of the MMPI studies would be more likely to establish convincingly the prognostic value of the test. It is, therefore, disappointing to discover the large amount of disagreement in the MMPI findings. In about one third of the studies high scores on specific scales were thought to signify good outcome for psychoses, in another third high scores indicated poor outcome, and in the remaining third no prognostic value was found. Thus, high scores on the schizophrenia scale, the one most frequently cited, were thought by Harris, Bowman, and Simon (26), Hales and Simon (21), and Carp (8) to be favorable for outcome in the psychoses, while Harris (24), Pearson (51), and Feldman (16) felt high schizophrenia scores were unfavorable. A similar situation existed with the other MMPI scales, high scores sometimes being claimed indicative of good outcome, but as often found to be correlated with poor outcome or of no prognostic value.

It is clear that such evidence is not very helpful in providing a basis for prediction. It is quite likely that some of the studies provide more valuable evidence than others, but unless statistical estimates of the validity of the findings are provided, it is difficult to make a distinction between the good and the poor studies. Perhaps of even greater importance
is the paucity of confirmatory studies with new samples. Such confirmation has been attempted in four instances with the MMPI, but only two of these attempts proved successful (16, 52). One of the others failed to specify the criteria of outcome (26), while the second claimed validation on the basis of insignificant data (51). Again, the most accurate evaluation of the prognostic efficacy of the MMPI must be: not proven.

Studies of the use of ability tests for prognosis also fail to yield a consistent basis for prediction of outcome. Of 23 reports of the prognostic value of various psychomotor tests with psychotics, 10 indicated that favorable outcome correlated with efficient performance before therapy (6, 9, 20, 33, 38, 42, 49, 67, 77, 83). On the other hand, five reports concluded that inefficient performance was prognostically favorable (35, 43, 44, 48, 84), three reported either a curvilinear relationship or contrary trends under different conditions (36, 68, 79), and five found no relationship between intelligence tests and outcome (19, 26, 53, 60, 74). Such a summary of the results of the various ability tests may do injustice to the particular types of tests employed or special abilities studied. There is little evidence, however, that any particular test is consistently more effective for prognosis than another, or that the nature of the test can account for the divergent conclusions found in the literature. Again, the verdict is: not proven.

We have found, then, that with each of the three techniques, the Rorschach, the MMPI, and tests of ability, there is a large amount of disagreement concerning both the signs to be regarded as prognostic and the direction of the outcome to be predicted from any particular sign.

To some extent specific findings seem to be attributable either to the diagnosis of the patients or apparently to the type of treatment. For example, Malamud and others (43, 44) have reported that involuntary psychotics of estimated inferior intelligence have better prognoses than do those of above average intelligence. Studies by Schnack, Shakow, and Lively (68) and by Zubin and Thompson (83) have indicated that different prognostic signs may apply to patients given insulin than to patients given metrazol therapy. Numerous investigators (59, 68, 85) have also pointed out that the duration of the follow-up period may also play a role in determining the significance of the prognostic signs.

It is difficult to try to account for the divergent conclusions of the prognostic studies on the basis of any general characteristics, because relatively few investigators specified the important conditions which characterized their experiments. Thus, of 28 prognostic studies with the Rorschach, 19 (all but 20, 22, 38, 40, 47, 56, 59, 66, 76) did not specify the duration of illness of the patients and seven (30, 34, 54, 55, 63, 67, 75) failed to state the time of follow-up. Of 10 MMPI prognostic studies (8, 15, 16, 21, 24, 26, 50, 51, 52, 69), five reported that the prognoses applied to the broad diagnostic category of "psychiatric patients" (8, 15, 16, 24, 26, 52), only about one third specified the duration of illness of their patients (8, 21, 26), and about one third did not indicate the duration of the follow-up period (8, 24, 69).

None of these three factors, diagnosis, type of therapy, or duration of follow-up, seems to be able to account for all of the discrepancies in the literature. Apparently some other factor must be found to account for these contrary trends. The first question that needs to be answered is: Do the more efficiently functioning or the less efficiently functioning patients have the better outcome or is there no relationship at all?

In several instances in which inefficient pre-treatment test performance was correlated with good outcome, it was observed that the majority of the patients under study were chronically ill (26, 35, 43, 44, 48, 84). It was thought that, since duration of illness is well known to be an important factor in prognosis, it is highly probable that radically different prognostic signs would apply to patients differing in duration of illness. To test this hypothesis, data were gathered for both chronically ill psychotics (long duration of illness) and acutely ill psychotics (short duration of illness) on the Complex Reaction Time Test (79), a test which in essence provides a measure of the patient's ability to react quickly and appropriately to complex tasks presented continuously.

Comparisons of the relationships between initial CRT scores and final outcome status revealed divergent trends; initially, the im
proved among the acutely ill patients had the higher scores while the improved among the chronically ill had the lower scores. The divergence between these relationships was found to be significant at the .05 level when tested by analysis of variance. Even more striking, individual analyses revealed that a cut-off score of 30 was significantly differential for both groups of patients, but in opposite directions.

An individual analysis of the scores revealed that in the chronically ill, none of those who later left the hospital had scores of 30 or more, whereas half of those who were still in the hospital at the time of the follow-up had scores above this critical point. In the acutely ill, this same critical score was found significantly differential in the reverse order. In this case, all of the improved fell above the critical score of 30, while of the unimproved about one third fell below that point.

This study indicates that a possible explanation of the contradictions in the literature dealing with the use of psychological tests for prognosis may be the neglect of the factor of chronicity. Because relatively few studies have made a point of specifying this factor, it is hard to determine whether or not this variable can satisfactorily account for all the observed differences in results. But it can be seen to be important in several cases, especially for those studies using ability tests or the MMPI.

Apparently most of the studies employed either groups of acutely ill psychotics or groups of psychotics of various degrees of chronicity. Since patients of longer duration typically have poorer prognoses and perform less efficiently than acutely ill patients, it is to be expected that a positive relationship between pretherapy performance and outcome would be found when the group under study consists of a mixture of chronically and acutely ill patients. Many of the studies reporting an inverse relationship between performance level and outcome may have employed chronically ill patients only. The basis for the negative correlation between test score and outcome in such patients will be discussed later.

The findings that different prognostic signs apply to metrazol and insulin therapy (68, 83) may also be accounted for on this basis, since the more severely ill patients were more likely to be treated with metrazol and the less severely ill with insulin. It must be remembered, however, that all of the conflicting conclusions found in the literature cannot be explained on this basis alone, especially since our studies have dealt only with psychotics. Furthermore, it is not to be expected that all conflicts would be resolved with the control of merely this one factor, since other factors are undoubtedly of importance even if we have not as yet identified them and their contributions to prognosis.
There is little difficulty in accounting for the prognostic findings with acutely ill psychotics since it is to be expected that patients capable of efficient performance would tend to have the better outcome. For the chronically ill, however, it is contrary to common sense to predict that psychotics in whom the deviation from normal efficiency is relatively great should have better outcomes than do psychotics less altered in performance ability by the disease process. Consequently, it was desired to study more thoroughly the 48 chronically ill Columbia-Greystone patients (45) to verify the prognostic value of inefficient performance, and, if possible, to specify more clearly the nature and extent of this inefficiency (85).

Of the 44 pretreatment measures available, it was found that five differentiated between criterion outcome groups of 38 patients who had been cooperative enough to yield representative data on more than a third of the tests. In each case inefficient performance presaged favorable outcome. In general, it appeared that on conceptual tests those patients who subsequently improved functioned less well than those who did not improve, but on perceptual tests the former functioned somewhat better.

Type of therapy was unrelated to either outcome or prognostic indices, but, interestingly, was related to the psychiatric ratings of prognosis made at the time of testing. Patients who were expected to improve were found significantly more often than the others to have had brain operations, either immediately or during the follow-up period. The psychiatric prognostic ratings were found to be significantly related to outcome when all the patients were considered, regardless of testability, but were not found to be related in the cooperative cases alone.

This brings us to a discussion of the 10 uncooperative patients. All but one of these patients had an unfavorable outcome and a diagnosis of hebephrenic schizophrenia. (This one relation was, incidentally, the only connection found between diagnosis and outcome.) Apparently, it was on the basis of these deteriorated cases that the psychiatric prognoses achieved their relation with outcome in the entire group. Furthermore, these patients are apparently even less efficient than are those who later improve. We are reminded of Langfeldt's (36) conclusion that both clever patients and intellectually debilitated patients have catastrophic outcomes. Apparently the same situation exists in our study. Chronic patients who performed relatively efficiently and those from whom no representative performance could be elicited had remained continuously in, or had been permanently returned to, the hospital during the more than three years since testing. The hypothesized relationship between levels of performance and the

---

**Fig. 2. Hypothetical Relation between Pretreatment or Initial Psychological Test Performance and Outcome in Early and Chronic Mental Patients**
stages of chronicity can best be shown graphically.

Most patients may be expected to function somewhat below the normal in psychological tests, this decrement increasing with the chronicity of illness. In general, acutely ill patients function better than the more chronically ill and have better prognoses. Within the chronic group the better functioning seem to have rather poor prognoses, those of intermediate functioning have relatively good prognoses, and those who seem deteriorated have very poor prognoses.

That even such global techniques as the Rorschach might prove effective as prognostic agents when the protocols they afford are investigated systematically and objectively without the restrictions imposed by orthodox scoring and interpretation is demonstrated by McCall's study (40) of the preoperative protocols of the Columbia-Greystone patients (45). Utilizing a system of psychometric scales for scoring the responses (80), he found that of the 35 scales which were applied, five showed a significant prognostic value in retrospect at the .05 level. Two scales in which high scores might be regarded as reflecting perceptual clarity were positively correlated with outcome in a four-year follow-up (surface color and perception of plant life); three scales in which high scores might be regarded as reflecting conceptual clarity were negatively related to outcome (dehumanization tendency, popularity of percept, reaction time). On the basis of this retrospective study it may be tentatively concluded that the perceptually clear but conceptually confused patients had a better prognosis than the perceptually confused but conceptually clear patients. It should also be noted that not one of the orthodox scoring categories proved to be prognostic, though some derivatives of the orthodox system were found to have prognostic value when scaled psychometrically.

**SUMMARY**

A review of the literature on the use of psychological tests for prognosis revealed an extreme amount of variation and contradiction in the indices considered prognostic. In an attempt to partially explain these contradictions, the factor of chronicity was investigated in psychotics. The prognostic index for chronically ill psychotics was found to be the opposite of that for acutely ill psychotics on the Complex Reaction Time Test. Further analysis of the preoperative data of the chronically ill psychotics revealed a certain amount of generality for the finding that those who functioned well on psychological tests had a poor outcome. This prognosis applied regardless of therapy, but was limited to the cooperative patients. The uncooperative or deteriorated patients were almost always diagnosed as hebephrenic schizophrenics and had a very poor expectancy of remission. There was, then, a curvilinear relationship between performance and outcome in the chronic patients, those performing best and those unable to cooperate having a less favorable outcome than patients of low-level efficiency.

It is important to emphasize that these findings have been exploratory and require validation from confirmatory studies before confidence can be placed in the conclusions. The fact that we have found support in the literature for our findings is of relatively little importance since so many different claims have been made that almost any hypothesis can find considerable agreement. It is, then, essential to confirm or refute our hypotheses with additional patients in whom more factors can be controlled than was the case in our studies. This need for confirmation applies to most of the studies of psychological prognosis.

Although it may be premature to speculate on the implications of our hypotheses concerning different prognostic signs for patients differing in chronicity, it is not too soon to point out some of the criteria for a good prognostic study. First, it is important to control and specify the conditions of the study. Among those which may be important are: (a) diagnosis, (b) type of therapy, (c) duration of illness, (d) stage of disease, (e) time of follow-up, and (f) opportunities for, and criteria of, outcome. Second, what is most needed is confirmation or refutation of existing claims of prognostic efficiency rather than new claims. Perhaps studies designed to test the range of conditions within which particular signs are prognostic would most satisfactorily establish the validity of previous claims. Last, the evaluation of the results should be statistically
acceptable, in contrast to the majority of present-day reports of prognostic psychological tests.

When investigators begin to apply some of these criteria, it may be possible to place some degree of confidence in the reports of prognostic efficacy for psychological tests.

REFERENCES


42. Malamud, W., & Rendell, N. Course and prognosis in schizophrenia. *Amer. J. Psychiatr.*, 1939, 95, 1039–1057.


69. Schreibfied, W. Changes in responses to the Minnesota Multiphasic Inventory following certain therapies. *Psychol. Monogr.*, 1950, 64, No. 5 (Whole No. 311).


71. Siegel, S. M. Personality factors in psychotherapeutic improvement as identified by the Rorschach test. *Amer. Psychologist*, 1951, 6, 341–342. (Abstract)


Received February 12, 1953.