PSYCHOLOGICAL MEASURES AS USED TO PREDICT PSYCHIATRIC IMPROVEMENT AND TO ASSESS BEHAVIOURAL CHANGES FOLLOWING PREFRONTAL LOBOTOMY*

By

PAUL McREYNOLDS

and

MARIAN WEIDE

Veterans Administration Hospital, Palo Alto, California

I. Introduction

The advent of the various tranquillizing drugs during the past few years and their increasing use in mental hospitals have had the effect of drastically decreasing the number of lobotomies performed. At our hospital, e.g., the last lobotomy was performed in November, 1954; previous to that time 131 such operations had been done in seven years. To all intents and purposes lobotomy is a therapeutic procedure no longer used for psychiatric purposes. This may change, of course—it may be decided that for certain patients a lobotomy is still the treatment of choice. Or, it may be that in the years to come lobotomy will be of interest chiefly to medical historians.

In any event, it seems desirable that we assess, as well as we can, the effects of the operation. Such information—quite aside from indicating possible therapeutic uses of lobotomy—may prove of theoretical value in furthering our understanding of brain functioning, and of heuristic value in our evaluations of other therapies.

The investigation to be reported here was carried out at the Palo Alto Veterans Administration Hospital from 1951 to 1955. It consisted, in the main, of the administration of certain psychological procedures to operated patients, both before and at intervals following operations; of the comparison of the obtained measures to certain available criterion data; and of the evaluation of changes in these measures over time.

* From the Veterans Administration Hospital, Palo Alto. The test battery used in this study was selected and/or developed by the senior author; the data were tabulated and analysed by the junior author; and both authors are responsible for this report. A number of other persons also participated extensively in this study. The patients were examined primarily by psychology trainees under the supervision of Dr. Richard C. Hamister. Mary Ann Swanson contributed greatly in the scoring and collection of data. The authors are particularly indebted to Dr. Kenneth P. Jones, who with the assistance of a large number of ward physicians, collected the criterion data. These data were collected for another study (7) but were generously made available to us. The physicians who assisted Dr. Jones were: Drs. G. Altbach, Wallace G. Beckman, Franklin C. Cassidy, Abram Dansky, Harry Elkins, Emerson Hiler, Genevieve Knupfer, Alan J. Kringel, William R. Meadows, Harold Mikkelsen, L. Mitchell, Robert S. Mowry, Judity Musladin, R. G. St. Pierre, David F. Shupp, John O. Smith, Leo M. Traub, and John A. Withrow. Dr. Robert L. McFarland had much to do with setting up the criterion rating system. Dr. Alexander Simon made a number of helpful suggestions during the planning phase of the study. Drs. Glenn Brackbill, John Daily, Richard Hamister, and S. D. Schultz generously acted as judges of Rorschach protocols.

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

II. General Plan of the Study

1. *The Lobotomy Programme at Palo Alto*

The lobotomy programme was inaugurated here in June, 1947. The last operation was performed in November, 1954. During this time 131 patients were operated—all but four having bilateral operations. Six patients died before follow-up criterion data had been collected. If these deaths are eliminated from consideration there remain 125 patients. Criterion data were gathered for this sample of 125 in November, 1955. The effective closing date of this study is thus November, 1955, which is one year after the date of the last operation.

At this hospital lobotomy was considered a treatment of last resort and was used only with patients who were extremely ill, and who had not improved under all other available forms of therapeutic intervention. The hospital was quite conservative in ordering the operation and routinely insisted that all other therapies be tried and found ineffective before a lobotomy was undertaken. Consequently, the operated patients comprised a sample of extremely ill and chronic subjects, such that almost any degree of overall improvement, had the operation not been performed, would have been unlikely.

2. *The Criterion of Improvement*

In order to evaluate the effectiveness of lobotomy it was necessary to have a measure of improvement after the operation. The improvement criterion used at this hospital was adapted from that used in the Columbia-Greystone (12) study. A modification of this scale first was used at this hospital in a study by Becker and McFarland (1) in their development of a lobotomy prognosis scale. In a report made in 1956, Jones (7), reporting on general lobotomy follow-up data, used a scale almost exactly like that used in the Columbia-Greystone study. As used by Jones the scale consisted of 5 steps, as follows:

0=Minimal or no improvement, patient on disturbed ward.
1=Patient on moderately disturbed ward, no ground privileges.
2=Patient has ground privileges.
3=Patient at home, under supervision.
4=Patient at home and working or capable of working.

Since the great majority of patients given lobotomies at this hospital were from disturbed wards*, the placement of patients on the above scale at some period after the operation would give a good approximation of the degree of improvement attained by those patients. Dr. Kenneth P. Jones, in November, 1955, obtained data to rate each of the 125 lobotomized patients on the above criterion. These data were furnished by the various ward physicians who had lobotomized patients under their care. On the basis of the reports Jones, who was the psychiatrist in charge of the ward to which most post-leucotomy patients were assigned, prepared a report (7) of the clinical course of the lobotomized patients.

Of the 125 patients, 22 obtained ratings of 0; 43 ratings of 1; 36 ratings of 2; 12 ratings of 3; and 12 ratings of 4. These distributions are presented graphically in Figure 1. It is to be noted that all of the 125 patients were rated at the same time—November, 1955. For some patients this was seven or eight years after the operation, for one patient as little as one year after operation.

* The number of patients from other wards was so small as not to merit separate consideration.

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

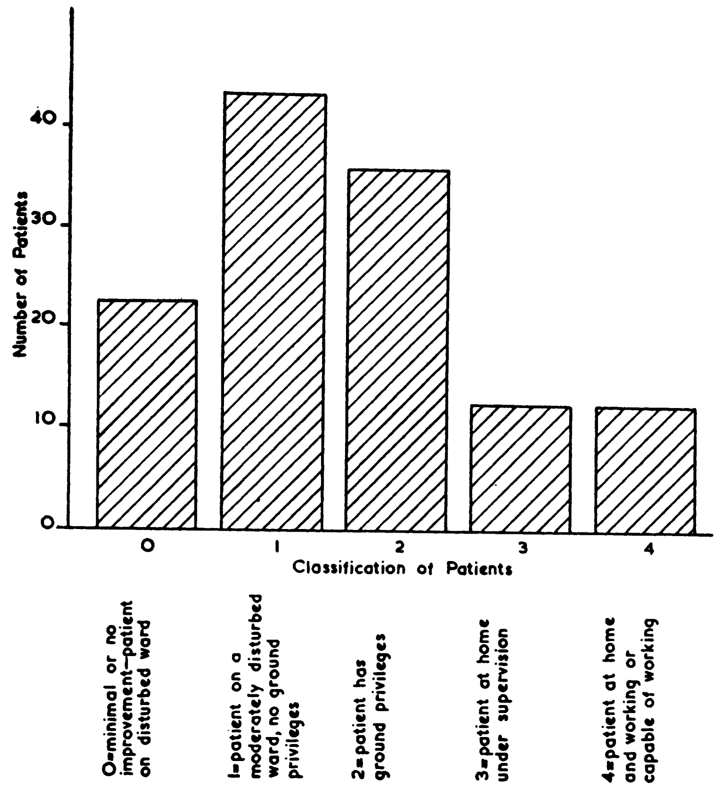

FIG. 1.—Classification of lobotomized patients as of November, 1955*.

* All operated patients were classified on the above scale November, 1955, regardless of when they received lobotomies. At least one year had elapsed between operation and rating for all patients.

It would have been preferable, for analysis and interpretative purposes, if the ratings had been obtained the same number of months or years after operation for each patient. It was not possible to do this, however, the fact that the criterion ratings for the different patients are based on different post-operative intervals, while unfortunate from the standpoint of experimental design, is probably not overly significant*, since it is generally believed that such changes as may follow a lobotomy occur predominantly within the first post-operative year.

These ratings, as obtained by Jones (7) in November, 1955, provided the basic criterion data for the study to be described in this paper.

3. *The Lobotomy Psychology Testing Programme*

The research lobotomy testing programme was begun in May, 1951, and terminated in April, 1955. The general plan of the programme was to administer, to each patient selected for lobotomy by the Hospital Lobotomy Committee, certain prescribed tests. These tests were given before the operation, 10 days following the operation, and at intervals of 5 weeks, 3 months, 6 months, and

* The correlation coefficient between the criterion ratings and the time since operation is equal to ·10, which is not significantly different from zero. This indicates that there is no significant over-all change in ratings after the first year following lobotomy.

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

1 year after the operation. The prescribed test battery was somewhat different for each of the periods. Table I presents the names of the tests involved, and the

TABLE I

*Schedule for Psychological Testing of Lobotomy Patients**

Tests	Time					
	Pre-operative	Post-operative				
		10th Day	5th Week	3rd Month	6th Month	12th Month
Rorschach	×					×
Draw-a-Person	×		×			×
Modified Porteus Maze	×			×		
Bender-Gestalt Cards 4 and 7	×	×		×		×
Concept Evaluation Technique	×	×			×	×
Colour Naming	×	×	×	×	×	×
Digit Symbol	×	×	×	×	×	×
Object Memory	×	×		×	×	×
Wechsler-Bellevue Vocabulary				×		
Sentence Completion	×		×		×	×
Motor Inhibition	×	×	×		×	×

* These tests are described in III 2 below. As already noted it was often impossible actually to obtain test data according to the above schedule. The schedule is best interpreted, therefore, as indicative of when attempts were made to obtain certain test data.

intervals at which they were to be given. During the period of the study 73 cases were assigned for research psychological testing. Testing was done primarily by psychology trainees and to the extent possible each patient was tested throughout by the same examiner. Of the 73 patients assigned for psychological testing, many were unable to take certain of the tests or, for various administrative reasons such as illness, trial visits and the like, were not tested at a given scheduled period. Thirty-four patients, e.g., were not given any of the pre-operative tests. The number of patients completing a given test at a given period ranged from 23 to 56.

The *purposes* of the lobotomy testing programme were to obtain additional information on the following two questions:

1. How useful are psychological tests in predicting which patients will benefit from lobotomy? and
2. What are the changes in mental functioning over time following lobotomy?

Obviously both of these questions are of such size that we could not hope to answer them definitively. Rather, our hope was to make a significant addition to a literature already rather voluminous (5, 6, 8, 12, 14, 16, 17, 18, 19).

In choosing the tests to be included in the battery we were guided by several considerations: (1) the procedures should be relatively brief and should be easily administered—since many different examiners would be involved; (2) as much as possible the procedures should be amenable to the most regressed of patients, since a great many of the patients to be operated were, in terms of standard test batteries, "untestable"; and (3) the procedures should have some relevance—either in terms of previous studies or on theoretical grounds—to the rationale of lobotomy.

Since we could include only a limited number of tests, no attempt was made to be comprehensive in sampling psychological functions. We were not attempting to develop a "better" battery than other investigators and did not select tests to comprise the best possible battery. Some of the tests were chosen because they are commonly used psychological instruments (e.g. Rorschach); some were included because there is evidence that they are related to the effects

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

of lobotomy (e.g. Motor Inhibition (3, 16)); some were included because it was believed that almost any co-operative patient, no matter how regressed, could take them (e.g. Colour Naming); and one (Concept Evaluation Technique, or CET) was included because of an hypothesis concerning its relation to the effects of lobotomy. In addition to the actual tests taken each patient was asked to help plan the order of the tests: this was to obtain data with which to evaluate the common hypothesis that lobotomy has a detrimental effect on the patient's foresight and planning ability. Each patient was also asked to estimate how long he had been on his ward, how long in the hospital, how much time the testing required, and the number of tests taken.

The nature of the tests used, and the constitution of the samples studied with them will be described in Part III.

4. *The Behavioural Rating Scale*

In order to obtain data concerning ward behaviour of the patients studied, the Hospital Adjustment Scale (4, 9) was used. Unfortunately the use of this scale did not commence until some time after the beginning of the study and, further, it was not possible to obtain periodic HAS ratings on all Ss even after use of the Scale was undertaken. Consequently, an insufficient number of HAS scores were obtained for these to be of use as an additional criterion in evaluating the predictive adequacy of the pre-operative psychological tests.

However, by comparing the HAS scores with the criterion of improvement ratings discussed in II 2 above it was possible to obtain information concerning the validity of the HAS, and this information is presented in V below. Further, it was possible to use HAS scores as a means of evaluating changes after lobotomy, and these findings are reported in IV 2 below.

III. Predictive Validity of the Psychological Tests

1. *Subjects*

As noted earlier, a total of 73 patients, including 69 males and 4 females, were seen one or more times for psychological testing. Because of the small number of females separate analyses were not carried out for each sex. Of the 73 patients, 14 had a criterion rating of 0; 23 of 1; 22 of 2; 8 of 3; and 6 of 4. Of the original group of 125 patients classified in the Jones study, all but two were diagnosed as schizophrenics. The ages of the 125 patients at the time of operation ranged from 20 to over 60 with the median age at 33*.

As already noted, the patients were so ill that it was often impossible to obtain adequate test data. Consequently, the number of patients on whom test data were available differed appreciably from test to test, and from testing period to testing period.

2. *The Predictive Validity of the Tests Against the Criterion of Improvement Ratings*

Due to the great differences in the scoring systems used for the various tests it was not possible to use one measure of comparison for all tests. Where the test scores are continuous variables which could be considered as coming

* These data were not tabulated separately for the 73 patients used in the psychological testing programme. It was assumed that this latter sample, selected only on the basis of date of operation, would have approximately the same characteristics as the 125 patient sample, i.e. homogeneity as to diagnostic classification and heterogeneity as to age. The age distributions would vary somewhat from test to test, and sufficient patients in each age group would not be available for any one test to make age a variable worth including in the analysis.

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

from approximately normally distributed populations, the Pearson product-moment correlation coefficient was calculated. The results for these tests are presented in Table II. The method of comparisons varied for the other tests.

A few brief comments will be made about the findings for each test procedure.

TABLE II

Correlations Between Certain Pre-operative Test Scores and Criterion Ratings

Test	N	r
Bender (Cards 4 and 7)	39	−·11
Concept Evaluation Technique:		
J score	28	−·42*
V score	23	−·24
E score	23	−·14
Digit Symbol	37	·12
Object Memory:		
Immediate	31	·06
Delayed	28	·19
Porteus Maze (modified scoring)	33	·23

* Significantly different from zero at the ·013 level (one-tailed test). None of the other coefficients reached statistical significance.

Bender-Gestalt. Cards 4 and 7 of this test were administered, and were scored by the Pascal-Suttell (13) scoring system. The correlation of the sum of these two cards with the criterion rating was not significantly different from zero.

Concept Evaluation Technique (10). The three scores obtained on this test, J, V, and E, were each correlated separately with the criterion. The only correlation of any significance is that for J; this is significant at the ·025 level for a two-tail test. Inasmuch as it was predicted in advance that this relationship should be negative, a one-tail test is not inappropriate—yielding an ·013 level of significance. The negative correlation indicates that a high J score, pre-operative, is contra-indicative for improvement from lobotomy. The theoretical interpretation of this result will be further discussed in III 4.

Colour Naming. The Colour Naming test given was that presented in Wells and Ruesch's Handbook (21). It was scored for both time and errors. The number of errors was mostly zero or very small, and hence not applicable to analysis. Since the distribution of time scores was highly skewed, a correlation coefficient was not calculated. The medians for each criterion group were computed: they showed no apparent trend.

Digit Symbol. This test was the digit symbol test of the Wechsler-Bellevue (20), either Form I or Form II. In most instances Form I was used for pre-operative testing. Weighted scores were correlated with criterion ratings, and the correlation was insignificantly different from zero.

Draw-a-Person. Most patients were asked to make only one drawing; this was usually of a man. For patients who had drawn both a man and a woman, only the drawing of the man was scored. The drawings were rated as "good", "fair", or "poor" by one of the authors (M.W.). "Poor" meant that the drawing was primitive or bizarre. An appropriate tetrachoric correlation coefficient was computed by grouping together the ratings of "good" and "fair", and by

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

grouping criterion classifications 0 and 1 against 2, 3, and 4. The obtained coefficient, ·19, is not significantly different from zero.

Motor Inhibition. This test was adapted from the Downey Temperament Scale (3), and had previously been used in lobotomy research (16). The patient was asked to write "United States of America" as slowly as possible but with the pencil always moving. He was told that some people could take as long as one-half hour to do this. The length of time taken by the patient was recorded. The distribution of these times was highly skewed, hence, only the medians for each classification criterion were computed. These failed to show any apparent trend.

Object Memory. Close to the beginning of the testing sessions the S was asked to identify the pictures of 10 objects (21, page 155) and then was asked to recall them immediately after the identification. Near the end of the testing session he was again asked to recall the objects from memory. The number of objects correctly recalled from these two trials were both correlated with the criterion ratings. Neither correlation was significantly different from zero.

Porteus Mazes (15) (*Modified Scoring*). The usual Porteus scoring method was not used. Instead each S was given a score corresponding to the highest age level at which he successfully completed a maze on any of the trials allowed. This variable did not correlate significantly with the criterion ratings.

Sentence Completion. A special sentence completion test was developed for this study. Five scores were obtained for this test: (*a*) the number of sentences attempted; (*b*) the percentage of these which were grammatically correct; (*c*) the percentage which showed perseverances in language or thought expressed; (*d*) the number out of a selected group of five sentences which showed a positive affect; and (*e*) the number out of these five which showed negative affect. Since there were no apparent trends when these data were compared with criterion data, detailed analyses were not performed.

Vocabulary (*Wechsler-Bellevue*). This test was not administered before the operation.

Planning. The Ss were rated as: (*a*) unwilling to help plan the order of testing; (*b*) willing but unable to help plan the order of testing; or (*c*) helped plan the order of testing. These data showed no trends when compared with criterion data; most patients in all criterion groups were unwilling to attempt planning the tests.

Estimates. Patients were asked to estimate: (*a*) the times required for testing, the time they had been on their ward, and in the hospital; and (*b*) the numbers of tests given. When these estimates, expressed as percentages of actual times or numbers, were compared with criterion ratings no trends were apparent, and no detailed analyses were performed.

Testability. It was felt that the testability or untestability of the patients before operation might itself be related to post-operative improvement. Inspection showed that approximately 50 per cent. of *all* criterion classifications were untestable before the operation, however; hence this dichotimization had no predictive value.

Rorschach. The Rorschach data were analysed in two ways, as follows:

(*a*) The total number of responses given and the breakdown of this number by determinants was considered. These results were converted to T-scores (2) for analysis. Figure 2 gives the psychograms of the mean scores for each classification group. The FM scores seemed to have a fairly widespread scatter; hence, a one-way analysis of variance was performed to test the

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

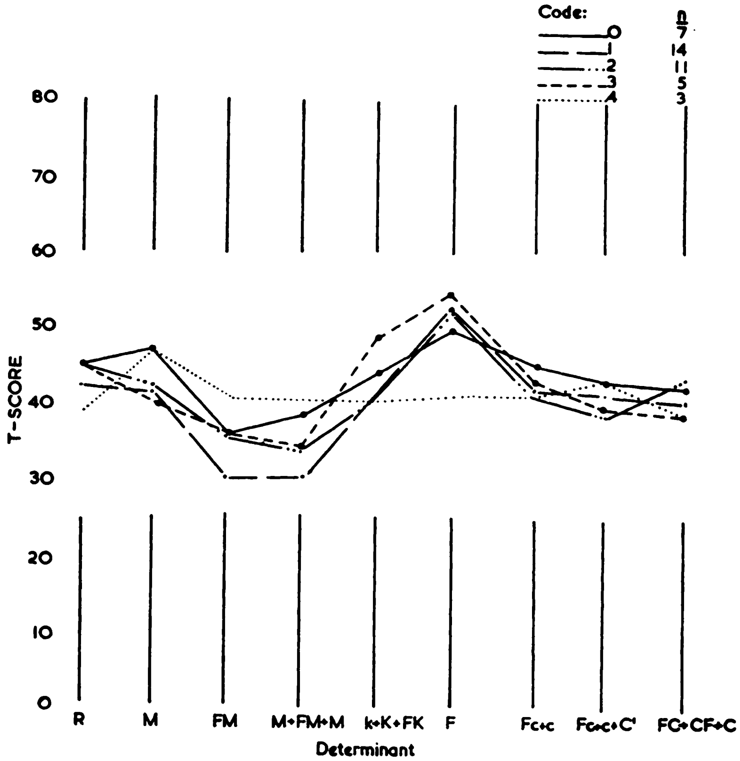

FIG. 2.—Pre-operative Rorschach psychograms of mean t-scores by classification groups.

differences among the means. The results were not significant, and accordingly no further analyses of scores were undertaken.

(*b*) Four experienced clinical psychologists, all of whom were familiar with both Rorschach procedures and lobotomy follow-ups, were given the available Rorschachs (N=38) for judgment. On the basis of these protocols they predicted the criterion rating each of the 38 patients would receive. These predictions were correlated with the actual ratings; none of the correlations approached significance. They ranged from —·01 to ·15.

3. *The Predictive Validity of the Tests Against the Criterion of Behavioural Ratings*

As mentioned before, the Hospital Adjustment Scale also was considered for use as a criterion of improvement. However, there were not a sufficient number of ratings made at any one period, either around a given date or a given time after operation, for this to be possible.

4. *Discussion of Findings Concerning the Predictive Utility of the Tests*

The findings of this study, as so far reported, are, with one exception to be discussed presently, negative. The utility of the instruments used to predict whether or not, and how much patients will improve after lobotomy is so little as to provide no justification for the inclusion of the instruments in any battery of tests designed to help select patients for operation. There are, of course, extenuating circumstances. For one thing, *the population studied was very*

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

homogeneous—all of the patients were very seriously ill and *all had passed through several screenings prior to being tested pre-operatively*. The point here is this: that in effect "the cards were stacked" against the tests from the beginning: the tests used might well have proved valid if the sample had been more heterogeneous—they might well have proved useful in screening patients to be considered for lobotomy. But the practical conclusion seems to be, nevertheless, that *those tests used in this study*, had they been used as selective criteria for determining which of the patients considered in this study should have been operated, would not have added materially to the adequacy of such selections.

The one exception to this generality is the J score of the CET (Rorschach Concept Evaluation Technique). It is interesting to note that this is the one test score in the whole battery for which a specific hypothesis was formulated in advance of the testing. This hypothesis was that J should correlate negatively with improvement after lobotomy. As noted in Table II this hypothesis was confirmed ($r = -\cdot 42$).

The CET (10) is a procedure in which the S is successively shown 50 areas on the Rorschach blots and asked, for each one, if it could be some specified object. Thus, item number one involves the presentation of blot 1 with the question "Could this be a bat?" The S simply answers Yes or No. The number of Yes answers he gives for the entire 50 presentations is counted, and this value is translated into a T-score. This T-score is referred to as J. The actual finding, therefore, was that those patients with initial low J-scores have a better chance of improving after lobotomy than those patients with initial high J-scores.

What is the meaning of this trend? And why was such an hypothesis formulated in the first place?

The J-score is assumed to measure the degree to which a S's standards of evaluation, or conceptualization, are strict and demanding—at the one extreme, or loose and non-rigorous—at the other extreme. To quote from a previous publication (10, page 66): . . . "The hypotheses for a low J-score would be that this individual tends to maintain, in his thinking, strict standards of evaluation, to have highly channelized thought patterns which would tend to revolve endlessly about a few subjects. He would tend to be ruminative, to take things seriously, to carefully consider all sides in making a decision. A high J-score, on the other hand, might suggest ramblings and tangential thought patterns, a tendency not to be ruminative or to take things seriously."

If this is what J measures, then what is the predicted effect of organic intrusions on these functions? Previous work and experience has suggested that patients with cerebral organic pathology tend to have high J-scores. It was therefore a rather straightforward expectation to suppose that the effect of lobotomy—which is a physiological intrusion—would tend to be to effect a raising of the J-score. This would mean that patients whose illness was manifested in a too-low J-score should tend to benefit from the operation, whereas patients with a too-high J-score should tend not to benefit from the operation. These effects would yield a negative correlation between J-scores and improvement.

There is one other interesting point to be made here, of a rather speculative kind. The above evidence would indicate that a patient's standards of conceptualization tend to become looser, less strict, after lobotomy. Now, there is also evidence—not from this study, but from the Columbia-Greystone study (12, p. 305, 311)—that following lobotomy there is often a marked decrease in level of anxiety. These two tendencies may be functionally related.

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

One of the authors of this report (P.M.R.) has proposed a theory (11) in which anxiety is equated with magnitude of unassimilated material. To bring about a reduction in level of anxiety an anxious person would have somehow to assimilate a sizeable portion of this material. Since assimilation is basically a process of conceptualization it follows that assimilation would be easier if one's standards of conceptualization were made less rigorous. Hence lobotomy —which appears to bring about a lessening of the standards of conceptualization (according to the evidence of this study)—could be expected to tend to effect a lowering in anxiety level (as indicated by the Columbia-Greystone data).

The finding that J-scores are negatively related to improvement after lobotomy can also be reconciled with the views of Jenkins (6, p. 87) on the nature of schizophrenia and lobotomy. Jenkins conceptualizes anxiety as being due to a morbid degree of resonance between the pre-frontal lobes and the diencephalon. Such a circuit of self-sustained neural activity would possibly cause a patient to become concretistic in his conceptualizing. If, then, the operation reduces the tendency toward "morbid resonance of neural circuits", as Jenkins postulates, it might also result in a lessening of the overly-rigorous conceptual standards.

IV. Psychological and Behavioural Changes Following Lobotomy

1. *Test Changes Over Time*

As indicated in Table I, the general plan of the study called for giving the same, or equivalent, test procedures to the same Ss at different pre-determined times following operation. It would seem that the resulting data should be particularly useful in assaying test changes following lobotomy. Such analysis should be especially pertinent to the question of what mental functions and capacities are improved or hindered by lobotomy and might possibly bear upon the greater question of which psychological functions are subserved by the frontal areas.

Actually, however—and unfortunately, the data available were far from ideal for analyses of this kind. Due to the facts that many patients, on any given test, were often untestable—or for various reasons of practicability were not tested, it happened that the number of cases on whom successive testings on the same tests were available was quite small. However, for most of the tests the number of patients tested at each scheduled testing period was sufficient to indicate trends. The point here is that while the Ns of Ss tested at each scheduled period were, for the most part, satisfactory, these Ns were not comprised of the same patients throughout. Consequently, the trend analysis data have to be interpreted as suggestive rather than as demonstrative. With this point in mind we may consider the general findings, which are presented graphically in Figure 3. For descriptions of the test procedure used see III 2.

Bender-Gestalt. As indicated in Figure 3, Bender scores (based on cards 4 and 7 only) show an increase from pre-operative testing to the testing ten days after operation, and then a general decline to the 12th month testing. Since a higher score denotes poorer performance, it is evident that performance was somewhat poorer immediately after surgery, and then gradually improved. It is not possible to evaluate the extent to which this change is due to practice.

Statistical Tests. Before going further several comments should be made about the place of statistical tests in evaluating changes over time. The Bender results, as summarized above, were not evaluated statistically in order to determine whether or not the changes over time (Fig. 3) were reasonably attributable

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

to chance. Similarly, such statistical tests were not made for any of the test changes over time. The reason for this is that—as noted earlier—the means obtained at each testing are based upon different but overlapping samples. This meant that valid significance values could not be attached to any statistic associated with the evaluation of trends, and, consequently, the results of statistical analyses would have been only suggestive. They would have given no more information than that obtained from an examination of the data in graphical form. If trend analysis had been restricted to those cases on whom successive testings were done, the number of cases would have been too small to justify much concern. Hence, it was deemed best to present the data for all cases available even though the trends must be interpreted as only suggestive.

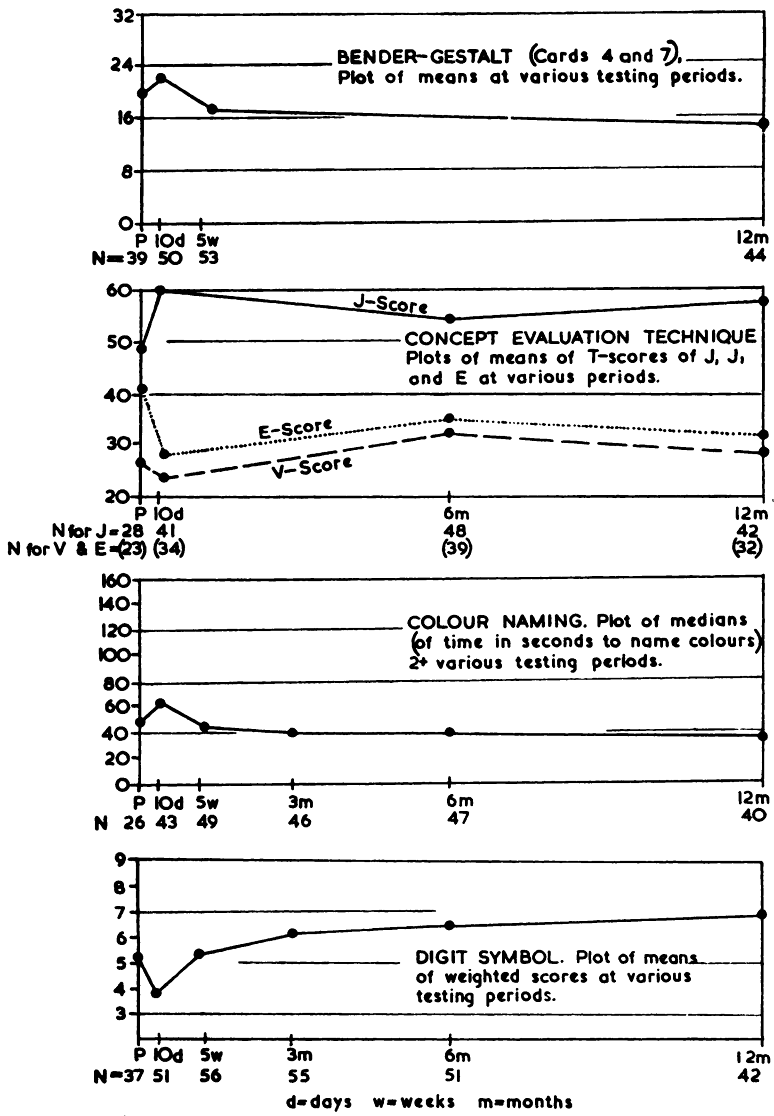

Figure continued on opposite page

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

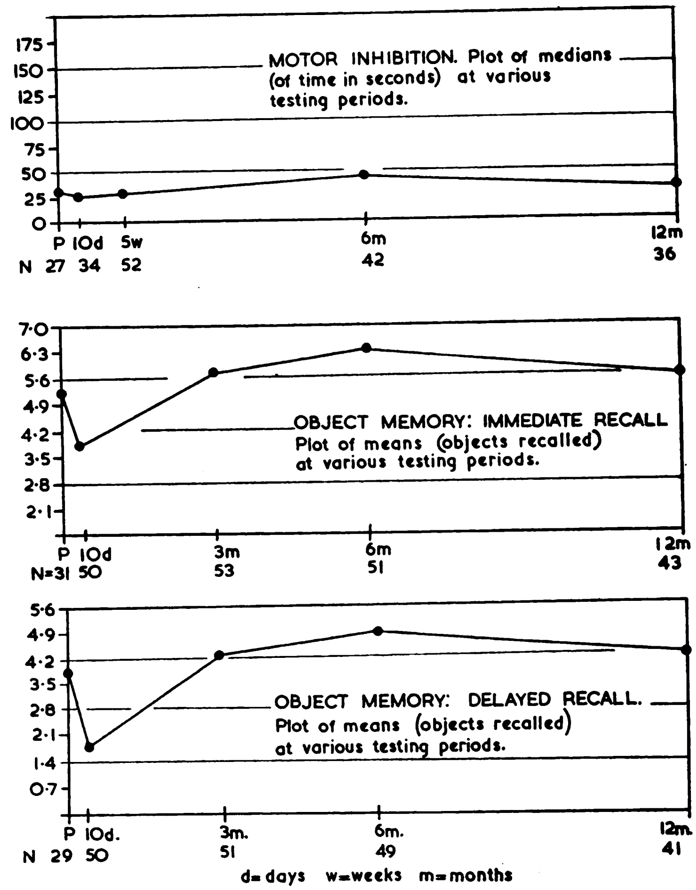

FIG. 3.—Plots of test scores against time after lobotomy*.

* Ordinates refer to test scores, abscissas to time. The extent of variability can be evaluated in terms of the range—in most instances the ordinates represent about one-half the range. The symbol P refers to Pre-operative scores.

The data are plotted on roughly comparable scales: for each test the range of values presented is about one-half the average range of test scores obtained.

Concept Evaluation Technique. As indicated in Figure 3 the J-score rises immediately after the operation and then tends to level off. The initial increase was more marked for patients who later obtained higher criterion ratings. The V-score showed an immediate decrease and then apparently an increase followed by a levelling off. The E-score showed a marked initial decline followed by a slow rise. The J-score trend is to be expected on the basis of previously discussed hypotheses; the V-score trend probably reflects an initial confusion followed by a minimal increase in degree of contact with reality; and the E-score trend can be interpreted as an initial loss of rigidity, followed by a trend toward increasing rigidity—which, however, never reaches the pre-operative level.

Colour Naming. Figure 3 presents the trend—in medians, which seemed more appropriate in this instance than means—for *time* required to name the colours. There is little change over the successive testings. Since the number of errors was heavily skewed toward zero, these data were not appropriate for analysis.

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

Digit Symbol. Here we note a drop immediately following operation, and then a gradual increase which appears to have levelled off by the end of one year. From inspection it appears that changes on this test are among the most dramatic found in this study. A suggestive implication is that the test improvement is due to an increased ability to concentrate, and, perhaps, an improved motor co-ordination.

Draw-a-Person. This test was given three times—pre-operatively, after five weeks, and after one year. The mean ratings assigned to the protocols are almost exactly the same for the three periods, and are not presented in Figure 3.

Motor Inhibition. Figure 3 shows the trend over time for this test, as plotted in medians, which in this instance were more appropriate than means. There is practically no change.

Object Memory. There were two parts to this test: the recall of objects *immediately* after seeing them, and the recall of objects after a *delay* corresponding to the rest of the testing session. As noted in Figure 3, both parts have similar curves. If one overlooks the decrement immediately following operation, then the curves are indicative of a general improvement which reaches a maximum at about 6 months, and then tends to decline.

No other test trends are included in Figure 3. In the case of the Porteus Mazes data were not available beyond the third month, and in the case of Sentence Completion none of the scores, for various reasons, were appropriate for examination of trends. There were no clearly discernible trends in terms of patients' ability to plan the tests, or in the accuracy of their estimates of time required for testings, and the like.

2. *Behavioural Rating Changes Over Time*

The Hospital Adjustment Scale (4, 9) was used—to the extent that clinical duties of ward staffs permitted—in order to evaluate actual behavioural changes shown by post-lobotomy patients, There were about 100 different patients rated at one time or another on the HAS, but, for various practical reasons, it was impossible to hold to a very precise schedule in terms of when ratings were obtained. Consequently, the groups of patients rated at any one interval period after surgery are to some extent different from the groups rated at other interval periods, and the number of cases on whom ratings were available also varies greatly from one period to another. Figure 4 presents the means for the various yearly intervals after lobotomy of the HAS total ratings. Statistical tests for the significance of trends revealed in these curves were not carried out. Such tests would have been inappropriate because of the facts just noted: that the patients involved, as well as the sample sizes, varied from interval to interval.

In the interpretation of these curves there probably is a certain degree of bias as the intervals get further from the date of operation—i.e. the later values may be somewhat too low since they obviously were based on patients still in the hospital, and did not include patients who had improved enough to be discharged. This number was small, however, and any such bias is minimal. It is to be noted that patients were included in the HAS ratings who had been operated several years before the psychology testing programme was undertaken. This made it possible to obtain ratings for periods up to eight years following surgery.

It is further to be noted that only one HAS score—the total score—is reported. Actually, the HAS includes also three sub-scales—concerning communication and interpersonal relations; care of self and social responsibility;

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

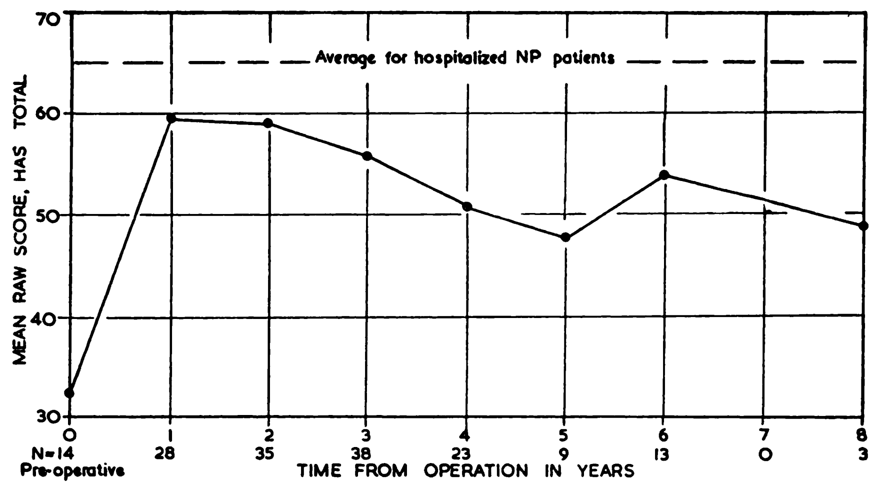

FIG. 4.—Mean HAS raw scores plotted against time after lobotomy*.

* The first year interval includes cases rated from 10½ to 15 months after surgery; the second year interval cases rated from 21 months to 30 months; the third year interval cases rated from 30 months to 42 months; and the same for years thereafter.

and work, activities, and recreation. Plots of these separate sub-scales against time after operation are not presented since in general they appeared to follow the same trend as does the total HAS score. The mean total raw score obtained by the 14 patients rated pre-operatively was 32·7. This is equivalent to the 21st percentile value for hospitalized psychiatric patients in general (9), and supports the view that the patients operated comprised an extremely ill group. Since the average raw score for hospitalized psychiatric patients (50th percentile) is 65, it appears from inspection of Figure 4 that the typical lobotomized patient did not attain a degree of improvement commensurate with the average of hospitalized psychiatric patients. Nevertheless, the trend indicated in Figure 4 does imply a rather marked general behavioural improvement following lobotomy.

3. *Discussion of Findings Concerning Changes Following Lobotomy*

Because exact statistical tests of the significance of changes in test scores were inappropriate, as indicated above, any conclusions which can be made concerning changes over time following lobotomy must be interpreted as being only tentative and suggestive. With this qualification, it seems possible to draw several interesting and pertinent generalizations from the data.

In the first place, it is noteworthy that most of the test results show a *decrement* in performance immediately after surgery—as reflected by testings 10 days after operation. This tendency no doubt is a function of the physiological trauma occasioned by the surgery itself: at this stage the patient is, primarily, an "organic" patient. This immediate effect appears to leave rather quickly, however, except in the case of the J-score on the CET, where it persists indefinitely up to the time of final measurement.

The test scores which appear to represent possibly real and substantial changes after lobotomy include, in addition to J on the CET, the Bender-Gestalt, Colour Naming, Digit Symbol, and Object Memory. It is difficult to evaluate the possible effects of practice on these tests, but it is noteworthy that

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

in each case the change was in the direction of improved performance. Further, the improved performance appears to have levelled off by the end of the first year post-operative. In most instances the point of highest performance occurs well before the end of the first year.

These general conclusions are similar to those reported by Scherer, Winne, Clancy and Baker (17, 18), who interpret their data to indicate that performance on tests measuring mental efficiency and organicity drop sharply during the first two-week post-operative period, then rise to reach a peak at about three months after surgery, and then gradually decline. Of the tests used in this study, the Bender-Gestalt, the CET, Object Memory, Colour Naming, and Digit Symbol all are probably somewhat sensitive to organic deficit. The test results in general suggest a greater mental efficiency and improved retentive functions after the immediate effects of the operation have disappeared. The one trend which cannot be fitted into this pattern is the J-score on the CET, which appears to reflect, rather, a persisting "organic-like" effect. As noted earlier, this trend may imply a lessened anxiety level. If so, the changes in retentive and attentive performances probably are due to the lessened anxiety rather than to any effects of the surgery on these functions *per se*. Indeed, we would be very hesitant to suppose that pre-frontal lobotomy has any *direct* beneficial effects on memory capacities, inasmuch as there is considerable evidence elsewhere that memory is hindered by physiological interventions in the frontal areas. Unfortunately, our test battery did not include any measures of anxiety: this, as seen in retrospect, was perhaps the most glaring inadequacy in the test battery used in this study.

The test instruments used, except for the Rorschach, were not particularly appropriate for evaluating any changes in overall personality structure. In the instance of the Rorschach test, as indicated in Figure 5, the indication is that no such changes occurred. This result, of course, may be due to the lack of validity

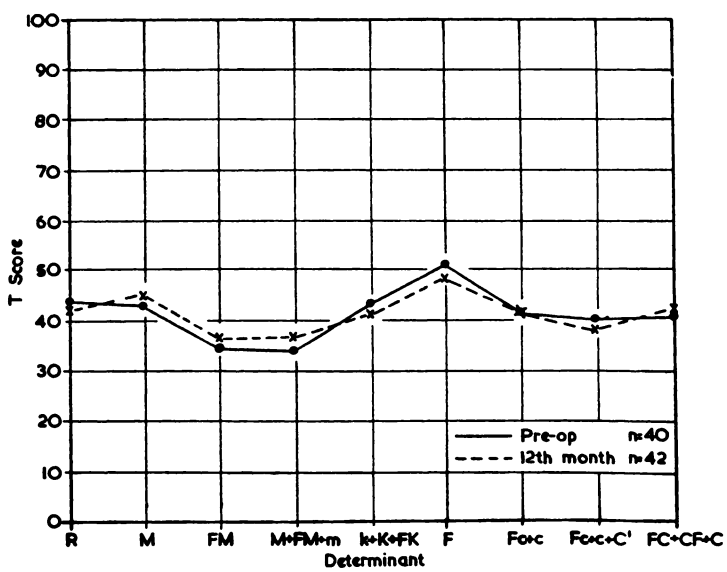

FIG. 5.—Rorschach psychograms of mean t-scores for pre-operative and one year post-operative testings.

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

of the Rorschach for this purpose. A more likely hypothesis, it seems to us, though, is that no overall personality changes did occur, i.e., that the effects of the operation were solely on the patients' mental efficiency—particularly retentiveness, attentiveness, and anxiety level—of the patients, without impinging directly upon their basic personality structures.

The HAS trend (Fig. 4) suggests an improved level of general, overall behavioural adjustment to the hospital ward living situations. In fact, the degree of improvement indicated is perhaps somewhat greater than would have been predicted on the basis of changes in test scores. If this is indeed the case, it may be due to changes in affect which were not measured by the tests used in this study.

V. The Relation of Hospital Adjustment Scale Scores and Criterion Ratings

1. *Findings and Discussion*

Both the HAS and criterion ratings based on the patients' ward status were used as indices of general change after lobotomy. It should be interesting therefore, to see to what extent these two measures are related. It will be recalled that HAS ratings were obtained periodically on the patients whereas the criterion ratings were obtained only once, in November, 1955. It was not possible, therefore, to select very many instances in which the HAS and the criterion ratings were obtained at exactly the same period.

By making some allowances, however, 85 cases were obtained on whom both HAS and criterion rating scores were available. These allowances were that the HAS closest to a criterion rating was matched with that rating, provided it was within at least one year of it. All HAS ratings used in this analysis except one were made at least one year after operation—i.e. after a point at which the scores had presumably reached considerable stability. This one exception was made only six months after operation.

For the 85 cases the HAS scores were correlated with the criterion ratings, yielding the coefficients presented in Table III.

Table III

Correlations of HAS Scores with Criterion Ratings

HAS Score	Correlation Coefficient	Level of Significance
Sub-scale I	·47	·01
Sub-scale II	·41	·01
Sub-scale III	·43	·01
Total	·52	·01

As noted in Table III all coefficients are positive and significant. Actually, the values probably are under-estimated somewhat, due to the fact that it was not possible to select data in such a way as exactly to match the times at which the two sets of measurements were obtained. The positive findings here increase one's confidence in the validity both of the HAS and the criterion ratings but have no other direct applicability to the lobotomy study.

VI. Summary and Conclusions

In May of 1951 a research project designed to evaluate prefrontal lobotomies by means of psychological instruments was inaugurated at the

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

Palo Alto Veterans Administration Hospital. Testing of lobotomy patients continued until April of 1955. During the intervening period 73 patients were examined.

The battery of psychological tests used included the Rorschach, the Rorschach Concept Evaluation Technique, the Draw-a-Person, a modified Porteus Maze, cards 4 and 7 of the Bender-Gestalt, Colour Naming, Digit Symbol, an Object Memory test, a Vocabulary test (Wechsler-Bellevue), a Sentence Completion test, and a test of Motor Inhibition. In addition behavioural ratings (Hospital Adjustment Scale) were obtained on many of the patients.

The testing programme was so arranged that certain of the tests were given at each of the following periods: pre-operative; and 10th day, 5th week, 3rd month, 6th month, and 1st year post-operative. The tests were administered by psychology trainees at this hospital. Many of the patients were unable to take some of the tests at certain prescribed testing periods: consequently the data available for analysis were not as great as would have been desirable, and systematic analyses of the data were not possible to as great an extent as could be wished. Nevertheless, the total amount of material gathered was very extensive, and certain generalizations can be drawn with considerable confidence.

The major purposes of the study were two: (1) to investigate the utility which those tests in the battery would have in predicting which of those patients operated would benefit maximally from the operation; and (2) to learn something about the course of changes in psychological functions which would occur following lobotomy.

With regard to the first of these questions—the predictive ability of the psychological tests—it was, of course, necessary to have some criterion of patient improvement against which to evaluate the tests. The major such criterion used was a rating based on the placement of the patients as of November, 1955. This rating varied from 0 (minimal or no improvement, patient on disturbed ward) to 4 (patient at home and working or capable of work).

When the test scores obtained by the patients pre-operatively were compared with the criterion ratings obtained by the patients in November of 1955 (this was at least one year post-operative for all patients) it was found that the relationship was statistically significant for only one of the tests (Rorschach Concept Evaluation Technique). That is, for the sample studied, all of the other tests used failed to show significant predictive validity as to which patients would benefit most from lobotomy. This finding should not be taken to mean that these tests are necessarily not useful in selecting patients for lobotomy, but rather that *after patients have already gone through careful screening for lobotomy* the use of these tests does not appear to add any significant predictability.

As to the second question—concerning changes in patients following lobotomy, as measured by the tests used in this study, the data indicated that most psychological functions measured tend to decline immediately following lobotomy—during which period the patient gives a rather typical "organic" picture—then to rise gradually with the optimal improvement apparently reached before the end of the first year post-operative. The improved functions appear to be chiefly in mental efficiency and retentive functions. It is also hypothesized that there is a significant decrement in level of anxiety, though our results did not bear directly upon this question. It is our hunch that the apparent improvements in retentive functions is in fact due to a lessening of anxiety rather than to any direct effect of the operations on retentive capacities.

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press

Our evidence indicates that prefrontal lobotomy did not bring about basic personality change. This tentative conclusion is based on the fact that Rorschach psychograms were not changed appreciably by operation.

The Hospital Adjustment Scale data indicated that, in general, the behavioural adjustment of operated patients—when considered as a group—improved rather dramatically during the first year post-operative, and then declined slowly thereafter. The improvement never reached the average for hospitalized neuropsychiatric patients in general, but at the same time the improvement, such as it was, was maintained to at least some degree for at least eight years post-operative.

The major findings of this study may be stated briefly as follows:

1. Scores of the one variable (J) of the Rorschach Concept Evaluation Technique given pre-operatively correlated significantly with the rated clinical status of the patients a year or more after operation; other test scored failed to show any significant correlations with the criterion.

2. After operation there was a tendency for test scores reflecting attentive and retentive capacities to decline, then to rise slowly to beyond the pre-operative level: these improved functions are interpreted as being due to a decrement in anxiety level, and

3. Behavioural adjustment scores (Hospital Adjustment Scale) tended to rise up to one year post-operative, and then to decline slowly.

References

1. Becker, W., and McFarland, R., "A lobotomy prognosis scale", 1954. Unpublished paper. Palo Alto, California: Veterans Administration Hospital.
2. Cass, W. A., "T-score conversion table for use with Rorschach Test scores", 1950. Unpublished research. Palo Alto, California: Veterans Administration Hospital.
3. Downey, June E., *The Will-temperament and its Testing*, 1925. Yonkers-on-Hudson: World Book Company.
4. Ferguson, J., McReynolds, P., and Ballachey, E., *The Hospital Adjustment Scale*, 1953. Stanford, California: Stanford University Press.
5. Greenblatt, M., and Solomon, H. (eds.), *Frontal Lobes and Schizophrenia*, 1953. New York: Springer.
6. Jenkins, R., Holsopple, J., and Lorr, M., "Effects of prefrontal lobotomy on patients with severe chronic schizophrenia", *Am. J. Psychiat.*, 1954, III, 84–90.
7. Jones, K. P., "A study of leucotomy data", 1956. Unpublished study. Palo Alto, California: Veterans Administration Hospital.
8. Jones, R., "Personality changes in psychotics following prefrontal lobotomy", *J. abnorm. soc. Psychol.*, 1949, **44**, 315–329.
9. McReynolds, P., and Ferguson, J., *Clinical Manual for the Hospital Adjustment Scale*, 1953. Stanford, California: Stanford University Press.
10. *Idem*, "The Rorschach Concept Evaluation Technique", *J. proj. techniques.*, 1954, **18**, 60–74.
11. *Idem*, "A restricted conceptualization of human anxiety and motivation", *Monograph Supplement to Psychol. Reports*, 1956, **6**, 293–312.
12. Mettler, F. (ed.), *Selective Partial Ablation of the Frontal Cortex* by Columbia-Greystone Associates, 1949. New York: Paul B. Hoeber, Inc.
13. Pascal, G. R., and Suttell, B. J., *The Bender-Gestalt Test, Quantification and Validity for Adults*, 1951. New York: Grune and Stratton.
14. Peters, H., "Traits related to improved adjustment of psychotics after lobotomy", *J. abnorm. soc. Psychol.*, 1947, **42**, 383–393.
15. Porteus, S. D., *The Porteus Maze Test and Intelligence*, 1950. Palo Alto: Pacific Books.
16. Robinson, Mary Frances, "What price lobotomy?" *J. abnorm. soc. Psychol.*, 1946, **41**, 421–437.
17. Scherer, I., Winne, J., Clancy, D., and Baker, R., "Psychological changes during the first year following prefrontal lobotomy", *Psychol. Monogr.*, 1953, **67**, Whole No. 357.
18. *Iidem*, and Baker, R., "Psychological changes over a three-year period following bilateral prefrontal lobotomy", *J. const. Psychol.*, 1955, **19**, 291–298.
19. Schrader, P., and Robinson, Mary Frances, "An evaluation of prefrontal lobotomy through ward behaviour", *J. abnorm. soc. Psychol.*, 1945, **40**, 61–70.
20. Wechsler, D., *The Measurement of Adult Intelligence*, 1938. Baltimore: William and Wilkins.
21. Wells, F. L., and Reusch, J., *Mental Examiners' Handbook*, 1945, Psychological Corp.

https://doi.org/10.1192/bjp.106.442.256 Published online by Cambridge University Press