CLINICAL DETECTION OF INTELLECTUAL DETERIORATION ASSOCIATED WITH BRAIN DAMAGE

DANO A. LELI

University of Alabama in Birmingham

SUSAN B. FILSKOV

University of South Florida

Leli and Filskov (1979) reported cross-validated classification accuracy that equalled 83% for a discriminant function derived on two measures of intellectual deterioration. This investigation made a preliminary assessment of the clinical utility of this function through a clinical-actuarial classification paradigm. Wechsler-Bellevue Intelligence Scale Form I protocols from 12 nonpsychotic nonimpaired and 12 cerebrally impaired individuals were used by experienced clinicians and predoctoral interns to identify the presence of intellectual deterioration associated with brain damage through their own clinical experience (Clinical Judgment condition) and, then, in conjunction with the discriminant function (Clinical-Actuarial condition). The classification accuracy from the discriminant function weights (Actuarial condition) and those from clinicians in the Clinical-Actuarial condition were statistically comparable and significantly above chance levels. These results indicate that the clinician who is assessing for the presence of intellectual deterioration associated with brain damage should rely heavily upon a valid actuarial index.

For the past 40 years, clinicians have been asked to determine the presence of intellectual deterioration associated with cerebral lesions. Various deterioration formulae and ratios (e.g., Allen, 1948; Hewson, 1949; Reynell, 1944; Wechsler, 1958) designed to aid the clinician in this task have not been empirically demonstrated to be useful (e.g., Anderson, 1950; Angers, 1958; Fisher, 1964; Kraus & Selecki, 1965; Reitan, 1959; Vogt & Heaton, 1977).

On the positive side, Leli and Filskov (1979) quantified the relationships of educational level (DED) and occupation (DOCC) with the Wechsler-Bellevue Full Scale IQ as two signs of intellectual deterioration. For DED, a Full Scale IQ greater than that expected from the *S*'s educational level was scored a 3. A score of 2 was given when the Full Scale IQ was comparable to that expected from educational level. A score of 1 was applied when a Full Scale IQ was lower than that expected from educational level. The identical scoring procedure was used for Full Scale IQ and occupation to derive DOCC. These two variables then were used as independent measures in a linear stepwise discriminant function, which was derived and cross-validated on cerebrally impaired and nonpsychotic nonimpaired adult *S*s. A subsequent cross-validation study with aged normal *S*s and individuals with dementia of the Alzheimer's type (Leli & Scott, 1982), who were administered the Wechsler Adult Intelligence Scale (WAIS), produced comparable classifications to those from the original cross-validation samples. While both sets of cross-validated classifications were high, these actuarial classification outcomes do not reflect the clinical utility of the weights from the discriminant function.

The basic assumption of this study was that clinical raters could detect accurately the presence or absence of brain impairment from a comparison of a global intelligence test score relative to premorbid estimates of intellectual functioning (educational level

[1]Reprint requests should be sent to Dano A. Leli, Ph.D., The Medical School, Department of Neurology, University of Alabama in Birmingham, Birmingham, Alabama 35294.

and past or current occupation). This general diagnostic method has been accepted clinically and empirically (e.g., Reitan, 1967) as a means of determining generalized functional loss as a result of medical and/or psychological disorders. Therefore, it was felt that an assessment of clinicians' abilities to use this classification method with and without a valid actuarial index would be heuristically interesting and clinically important. The purpose of this study was to make a preliminary assessment of the clinical utility of the classification information provided by the discriminant function weights developed by Leli and Filskov (1979) through a clinical-actuarial classification paradigm.

It was hypothesized that the classification information from the discriminant function weights would significantly improve inexperienced and experienced clinicians' abilities to detect the presence of intellectual deterioration associated with brain damage above that achieved without such information. Second, it was hypothesized that access to the classification information from the discriminant function weights would significantly improve inexperienced and experienced clinicians' detection of intellectual deterioration over that produced by the discriminant function weights. Third, it was hypothesized that inexperienced and experienced clinicians would not differ significantly in their ability to use the classification information from the discriminant function weights.

METHOD

Subjects

Two groups of clinical raters were formed. The experienced group consisted of five licensed clinical psychologists who were working in a medical school setting in which they frequently are asked to determine the presence or absence of neurological impairment of intellective abilities with the adult Wechsler scales. Their postdoctoral experience in neuropsychological assessment ranged from 1 to 23 years (M = 13.6 years). Three clinicians were ABPP Diplomates in Clinical Psychology. The inexperienced group consisted of five predoctoral interns in clinical psychology who had completed successfully a graduate-level intellectual assessment course and practicum.

Protocols

Wechsler-Bellevue Intelligence Scale Form I (W-B) protocols were obtained from brain-impaired individuals and nonpsychotic non-brain-impaired individuals who were administered the W-B alone or as part of the Halstead-Reitan Neuropsychological Test Battery (Halstead, 1947; Reitan, 1955). These individuals were assessed as inpatients, outpatients, and nonpatient volunteers from four university medical centers, a university outpatient psychological clinic, and the community. For the brain-damaged individuals, verification of cerebral impairment was obtained through medical tests, surgery, or autopsy. All of the nonimpaired individuals were considered to be normal in that they had a negative neurologic history and never had reported symptoms or exhibited demonstrable signs of cerebral pathology. Three of the nonimpaired individuals were considered to have psychological disorders (2 neurotics, 1 character disorder). These 3 individuals were considered to be comparable to the other nonimpaired individuals because they were not acutely disturbed or lacking motivation at the time of testing.

From this pool, 12 protocols from nonpsychotic, nonimpaired (9 male and 3 female) individuals and 12 protocols from brain impaired (6 male and 6 female) individuals with diffuse or lateralized lesions were selected. These 24 protocols were used previously as the initial cross-validation sample for the discriminant function and were the stimuli used in the Clinical Judgment and Clinical-Actuarial conditions described below. The size of this sample is consistent with the samples employed in other clinical neuropsychological classification studies (e.g., Goldberg, 1959; Goldstein, Deysach, & Kleinknecht, 1973). Fisher's Exact Probability Test indicated that the sexual composition of the two samples was not significantly different. For the nonimpaired subsample, the mean age was 51.42

(SD = 17.02), while the mean educational level was 13.33 (SD = 3.20). The mean age for the impaired subsample was 52.42 (SD = 15.68), while the mean educational level was 12.33 (SD = 2.93). t-test comparisons of these age and education means were not statistically significant ($p > .05$).

In both judgment conditions, each W-B protocol contained the person's age, sex, education, occupation, subtest and subscale raw and weighted scores, Verbal IQ, Performance IQ, Full Scale IQ, and Wechsler's (1958) Mental Deterioration Index (MDI). Also included in the bottom portion of the protocol were two boxes, labelled Yes (coded 1) and No (coded 2), which the rater marked in response to the following question, "Does this W-B protocol indicate the presence of intellectual deterioration associated with brain damage?"[2] Yoked to this diagnostic question was a 5-point confidence rating scale with Not Sure (coded 1) and Very Sure (coded 5) as the outer poles. A section for comments also appeared on the bottom of the protocols. In this section, the raters were asked to indicate which judgment method(s) they used in making their classification decision.

Included on the bottom portion of each protocol in the Clinical-Actuarial condition were the raw scores, unstandardized lambda weights, and the categorization score from the discriminant function. Each protocol also contained diagnostic cut-off scores (e.g., <1.00 Impaired, ≥ 1.00 Not Impaired) plus an interpretative statement (either Impaired or Nonimpaired) that indicated into which group the function placed the protocol. Each rater received written interpretive instructions as to the use of the discriminant function data with the protocols to be rated in the Clinical-Actuarial condition.

Procedure

Actuarial condition. The Actuarial condition consisted of the cross-validated classifications from the discriminant function weights previously reported by Leli and Filskov (1979). This condition was analogous to the actuarial conditions described by Goldstein et al. (1973) and Holt (1958) and the "mechanical composite" method described by Sawyer (1966).

Clinical judgment condition. In the Clinical Judgment condition, the raters were asked to classify independently the 24 protocols through their clinical experience and training. The raters were not informed as to the numbers of impaired and nonimpaired protocols. Prior to making their classifications, the interns were given a 15-minute presentation on the clinical utility of the MDI. Both groups of raters completed the protocols in the Clinical Judgment condition prior to rating the same protocols in the Clinical-Actuarial condition. This condition was analogous to the clinical judgment condition described by Goldstein et al. (1973), the "profile interpretation" method described by Sawyer (1966), and a weak form of "sophisticated clinical prediction" according to Holt (1958).

Clinical-actuarial condition. In the Clinical-Actuarial condition, the raters were asked to use the classification information from the discriminant function weights as part of their judgment strategies in independently classifying the same 24 protocols. With regard to the classification accuracy of the discriminant function weights, the raters were told that they "have previously demonstrated high predictive validity in identifying the presence or absence of intellectual deterioration associated with brain damage." This condition was analogous to the clinical-actuarial condition described by Goldstein et al. (1973), the "clinical synthesis" method described by Sawyer (1966), and a rigorous form of the "sophisticated clinical prediction" method described by Holt (1958).

The criterion utilized to establish classification accuracy was the rater's correct and incorrect classifications of the protocols into brain impaired and nonimpaired categories.

[2]The discriminant function weights were derived to answer the same classification question.

This was the same criterion used to establish the classification accuracy of the discriminant function weights. The correct classifications from the Clinical Judgment and Clinical-Actuarial conditions were tallied by the first author in a conservative fashion in that four of the five raters in each rater group had to agree on the correct classification of the protocol (Goldstein et al., 1973).

RESULTS

Table 1 depicts classification accuracy for the three classification conditions. The correct classifications (83.3%) from the discriminant function was significantly greater than chance ($p < .001$), as were the clinicians' correct classifications (75%) in the Clinical-Actuarial condition ($p < .01$).

TABLE 1

Classification Accuracy of the Three Classification Conditions

		Raters						
		Interns				Clinicians		
Classification condition		Correct	Incorrect	%		Correct	Incorrect	%
Clinical Judgment	NI	9	3		NI	6	6	
				62.5				58.3
	I	6	6		I	8	4	
Clinical Actuarial	NI	10	2		NI	11	1	
				66.5				75.0
	I	6	6		I	7	5	
Actuarial (Leli & Fiskov, 1979)	NI	11	1					
				83.3				
	I	9	3					

Note.—NI = Nonimpaired, I = Impaired.

TABLE 2

Chi-square Comparisons of the Classification Accuracy Produced from the Actuarial Condition and the Interns and Clinicians in the Clinical Judgment and Clinical-actuarial Conditions

	Interns		Clinicians	
Condition	Clinical Judgment	Clinical Actuarial	Clinical Judgment	Clinical Actuarial
Actuarial	3.789*	2.778*	4.941*	1.137
Clinical Actuarial				
Clinicians	.388	.101	1.333[a]	
Interns	.000[a]		.800	
Clinical Judgment				
Clinicians	.349			
Interns				

[a]Chi-square calculated from the McNemar Test for the Significance of Change.

*$p < .05$ ($df = 1$, 1-tailed test).

Table 2 presents chi-square tests, which were corrected for continuity, calculated on the correct and incorrect classifications from the discriminant function weights (Actuarial condition), and those produced by interns and clinicians in the Clinical Judgment and Clinical-Actuarial conditions. Looking from left to right across Table 2, actuarial classifications were significantly greater than the classifications produced by interns in the Clinical Judgment and Clinical-Actuarial conditions, and by clinicians in the Clinical Judgment condition. None of the remaining comparisons was statistically significant.

The correlations between classification accuracy and confidence ratings for the raters in the Clinical Judgment condition ranged from $r = -.15$ to .49 (mean $r = .19$) for interns and $r = -.25$ to .31 (mean $r = .13$) for clinicians. For interns, three of the five correlations ($r \geq .44$) were statistically significant ($p < .05$). None of the five correlations for clinicians was statistically significant. Correlations from the Clinical-Actuarial condition ranged from $r = .26$ to .77 (mean $r = .50$) for interns, and $r = .09$ to .55 (mean $r = .27$) for clinicians. Three of the five correlations ($r \geq .54$) were statistically significant ($p < .05$) for interns, while only one correlation ($r = .55$) was statistically significant ($p < .01$) for clinicians.

Discussion

The classification data depicted in Tables 1 and 2 indicate that the discriminant function weights were statistically superior to either inexperienced (interns) or experienced (clinicians) raters who were using only their clinical experience and training. These data also indicate that access to the classification information from the discriminant function weights failed to improve significantly the interns' classification accuracy. However, access to this information significantly closed the gap between correct classifications from the discriminant function weights and those produced by the clinicians. Access to the actuarial classification information also improved the clinicians' accuracy to above-chance levels. Similar outcomes were reported by Leli and Filskov (1981) in a more comprehensive clinical-actuarial neuropsychological classification study that used the W-B.

To clarify these findings, the classification accuracy of the individual raters was examined. In the Clinical Judgment condition, all of the clinicians and 4 of the 5 interns were less accurate than the discriminant function weights. One intern achieved parity with the discriminant function weights. However, this parity subsequently was lost in the Clinical-Actuarial condition due to an additional false negative error. Four clinicians and 3 students made a small nonsignificant improvement in their classification accuracy in the Clinical-Actuarial condition, which suggests that the additional actuarial data had a positive effect on their accuracy. Perhaps this small positive trend would have reached statistical significance had a larger number of protocols been given to the raters. Two clinicians produced classifications equal to that from the discriminant function weights, while a third clinician surpassed the actuarial condition by correctly classifying 21 (as opposed to 20) of the protocols.

The classification data from the raters both individually and collectively do not provide support for the first two hypotheses that clinical-actuarial classifications would be superior to those made in the Clinical Judgment and in the Actuarial conditions, respectively. However, the data do suggest that the clinicians, both individually and collectively, improved their classification accuracy to that achieved by the discriminant function weights when they had access to the actuarial classification data. This improvement seemed to be the result of a large reduction in the number of false positive errors (42%). It appears that the clinicians were more willing than the interns to incorporate the information from the discriminant function weights into their general classification strategies. It also suggests that the clinicians made effective use of additional valid actuarial data. This is in direct contrast with findings from other clinical assessment areas (e.g., Golden, 1964; Sines, 1959), in which increases in the amount of available information failed to improve raters' classification accuracy.

Comparison of the intern and clinician classifications from the Clinical Judgment and Clinical-Actuarial conditions indicates that classification accuracy was not affected differentially by clinical experience, training, or the classification information from the function weights. These findings are consistent with the third hypothesis that there would be no differences between the clinicians and the interns and are comparable to those reported in other clinical assessment areas (e.g., Oskamp, 1962).

The number of significant correlations between classification accuracy and decision confidence indicates that several of the interns tended to be more confident when they were classifying a protocol as nonimpaired in both judgment conditions. In contrast, only one clinician in the Clinical-Actuarial condition demonstrated this significant relationship. These data suggest that many of the interns may have had a classification bias toward minimizing false positive errors in both judgment conditions, while the clinicians did not exhibit any consistent bias. It is possible that the clinicians' lack of bias was due to their greater experience. However, such a speculation is in direct contrast to the clinical lore that clinicians become more biased and stereotyped in their clinical methods with increasing experience.

An explanation for this differential bias was sought unsuccessfully from an inspection of the number and type of classification strategies used by the raters. One clinician failed to report any of his strategies in either judgment condition. The remaining nine raters employed various combinations of the MDI, Verbal-Performance Scale differences, subscale pattern analysis, various pathognomonic signs, and the discriminant function data in the Clinical-Actuarial condition. Inspection did not reveal any consistent pattern as to type or number of strategies either within or across raters. The raters also rarely indicated which method they relied upon more heavily or how they combined the various methods to make their decisions. In retrospect, the raters might have provided such information had the instructions been more explicit.

The data from this investigation provide one tentative though important caveat to the clinician interested in using a global index of intelligence to detect intellectual deterioration due to brain damage. Correct classifications will be optimized when heavy reliance is placed on the classification information provided by a valid actuarial index of intellectual deterioration. The use of more traditional diagnostic methods is likely to result in only a chance level of classification accuracy with a maximation of false negative errors.

Additional studies obviously are needed before the weights from the discriminant function can be used routinely in the clinic. The discriminant function first must be cross-validated actuarially with the revision of the WAIS (WAIS-R) on larger neurological and psychiatric patient samples of various types. If these actuarial studies demonstrate high classification accuracy, a series of clinical-actuarial studies then would be needed to determine whether clinicians can effectively use the function weights. Such studies should systematically evaluate clinicians' ability to use the function weights both alone and in combination with various types of neuropsychological test data.

References

Allen, R. M. (1948). A note on the use of the Bellevue-Wechsler Mental Deterioration Index with brain injured patients. *Journal of Clinical Psychology, 4*, 415-418.

Anderson, A. L. (1950). The effect of laterality localization of brain damage on Wechsler-Bellevue indices of deterioration. *Journal of Clinical Psychology, 6*, 191-194.

Angers, W. P. (1958). A psychometric study of institutionalized epileptics on the Wechsler-Bellevue. *Journal of General Psychology, 58*, 225-247.

Fisher, G. M. (1964). The efficiency of the Hewson Ratios in diagnosing cerebral pathology. *Journal of Nervous and Mental Disease, 134*, 80-83.

GOLDBERG, L. R. (1959). The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender-Gestalt test. *Journal of Consulting Psychology, 23,* 25-33.

GOLDEN, M. (1964). Some effects of combining psychological tests on clinical inferences. *Journal of Consulting Psychology, 28,* 440-446.

GOLDSTEIN, S. G., DEYSACH, R. E., & KLEINKNECHT, R. A. (1973). Effect of experience and amount of information on identification of cerebral impairment. *Journal of Consulting and Clinical Psychology, 41,* 31-34.

HALSTEAD, W. C. (1947). *Brain and intelligence: A quantitative study of the frontal lobes.* Chicago: University of Chicago Press.

HEWSON, L. R. (1949). The Wechsler-Bellevue Scale and the Substitution Test as aids in neuropsychiatric diagnosis. *Journal of Nervous and Mental Diseases, 109,* 158-183, 246-266.

HOLT, R. R. (1958). Clinical and statistical prediction: A reformulation and some new data. *Journal of Abnormal and Social Psychology, 56,* 1-12.

KRAUS, J., & SELECKI, B. R. (1965). Brain atrophy and assessment of intellectual deterioration on the Wechsler Adult Intelligence Scale. *Journal of Nervous and Mental Diseases, 141,* 119-122.

LELI, D. A., & FILSKOV, S. B. (1979). The relationship of intelligence with education and occupation as signs of intellectual deterioration. *Journal of Consulting and Clinical Psychology, 47,* 702-707.

LELI, D. A., & FILSKOV, S. B. (1981). Clinical-actuarial detection and description of brain impairment with the W-B Form I. *Journal of Clinical Psychology, 37,* 623-629.

LELI, D. A., & SCOTT, L. H. (1982). Cross-validation of two indexes of intellectual deterioration on patients with Alzheimer's disease. *Journal of Consulting and Clinical Psychology, 50,* 468.

OSKAMP, S. (1962). The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monographs, 76* (28, Whole No. 547).

REITAN, R. M. (1955). Certain differential effects of left and right cerebral lesions in human adults. *Journal of Comparative and Physiological Psychology, 48,* 474-477.

REITAN, R. M. (1959). The comparative effects of brain damage on the Halstead Impairment Index and the Wechsler-Bellevue Scale. *Journal of Clinical Psychology, 15,* 281-285.

REITAN, R. M. (1967). Psychological assessment of deficits associated with brain lesions in *S*s with normal and subnormal intelligence. In J. L. Khanna (Ed.), *Brain damage and mental retardation: A psychological evaluation.* Springfield, IL: Charles C Thomas.

REYNELL, W. R. (1944). A psychometric method of determining intellectual loss following head injury. *Journal of Mental Science, 90,* 710-719.

SAWYER, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66,* 178-200.

SINES, L. K. (1959). The relative contribution of four kinds of data to accuracy in personality assessment. *Journal of Consulting Psychology, 23,* 483-492.

VOGT, A. T., & HEATON, R. K. (1977). Comparison of Wechsler Adult Intelligence Scale indices of cerebral dysfunction. *Perceptual and Motor Skills, 45,* 607-615.

WECHSLER, D. (1958). *The measurement and appraisal of adult intelligence* (3rd ed.). Baltimore: Williams & Wilkins.