RESEARCH ARTICLES

A Clinical Evaluation of the Neurobehavioral Cognitive Status Examination in a General Psychiatric Inpatient Population

C.J. Lamarre M.D.,[1] S.B. Patten M.D.[2]

[1]Professor, Department of Psychiatry, The University of Calgary, M-7 Bow Valley Centre, Calgary General Hospital, Calgary, Alberta

[2]Clinical Fellow, Alberta Heritage Foundation for Medical Research,Department of Community Health Sciences, The University of Calgary, Calgary, Alberta

The Neurobehavioral Cognitive Status Examination (NCSE) is a structured test of cognitive functioning. The NCSE assesses a broader range of cognitive functioning than the Mini-Mental State Examination (MMSE), but remains brief enough to be administered at the bedside in clinical settings. The purpose of this study was to assess the sensitivity, specificity, predictive value and reliability of the NCSE as a clinical case-finding instrument for DSM-III-R defined organic mental disorders in psychiatric inpatients. Validity was assessed by comparing the results of the test (interpreted as "pass" or "fail") to a blind clinical assessment by an experienced psychiatrist. The NCSE was found to have superior sensitivity to the MMSE (83% *versus* 43%), but inferior specificity (47% *versus* 97%). The low specificity resulted in a positive predictive value of only 24%. The NCSE had good test-retest reliability (Kappa = .69), but the inter-rater reliability was not as good (Kappa = 0.57). The NCSE was too non-specific to be used as a case-finding instrument for organic mental disorders. In conclusion, although clinicians may find the NCSE to be a valuable instrument for the assessment of cognitive function, it cannot be used as a screening or case-finding instrument for organic disorders among psychiatric inpatients.

Key Words: organic mental disorder, cognitive impairment, mental status examination, psychiatric patients.

INTRODUCTION

Over the past twenty years, a large number of structured cognitive function tests have been developed for clinical use. One potential clinical application of these tests is as screening or case-finding instruments for organic mental disorders. Our literature search uncovered forty publications concerned with ten instruments designed to measure cognitive function. The application of these instruments as screening or case-finding tools is emphasized in this literature by the frequent

use of sensitivity and specificity as measures of the validity of the instruments. The nature of this literature suggests that the primary value of these instruments is related to their ability to help detect or diagnose organic mental disorders. An alternate application of these tests is simply to use them as a measure of cognitive functioning, without reference to the relationship of specific test results to specific mental disorders. In this sense, the tests could be used as bedside alternatives to more extensive neuropsychological testing. One of the most recently developed instruments is the Neurobehavioral Cognitive Status Examination (NCSE) (Northern California Behavioral Group 1988). In a recent study, utilizing a sample of neurosurgical patients, this test demonstrated superior sensitivity to the Mini-Mental State Exami-

Table 1

Summary of DSM-III-R diagnoses

Organic mental disorders	n = 12
Mood disorders	n = 43
Schizophrenia	n = 9
Delusional disorder	n = 1
Psychotic disorders NEC	n = 3
Other	n = 2
Total	n = 70

nation (MMSE)(Schwann et al. 1987). However, in this study, specificity (which represents the probability of a negative test in subjects without the disease being tested for) could not be evaluated since all subjects had organic disease. Furthermore, the results cannot necessarily be generalized to clinical psychiatric settings.

For these reasons, we conducted a study designed to assess the validity of the NCSE in psychiatric patients. The study used a general hospital inpatient sample consisting of a series of admissions to an inpatient psychiatric unit at the Calgary General Hospital. The study was designed to assess the sensitivity, specificity, predictive value and reliability of the NCSE in the detection of clinically relevant organic mental disorders. For this reason, a clinical diagnostic assessment was used as the gold standard for the measurement of validity and reliability. For the purpose of the study, clinically relevant organic mental disorders were defined as those which fulfil of DSM-III-R criteria for an organic mental syndrome: delirium, dementia, organic amnestic syndrome, organic delusional syndrome, organic hallucinosis, organic mood syndrome, organic anxiety syndrome, organic personality syndrome, intoxication or withdrawal.

REVIEW OF THE LITERATURE

A large number of cognitive screening tests have been developed in the past three decades. Their development has been accompanied by an extensive literature studying and comparing the instruments (Lamarre and Patten 1991). Kahn et al (1960) produced the original instrument: "the Mental Status Questionnaire" in 1964. Blessed et al (1968) related morphological changes in the brain to test scores and diagnosis in their work developing The Blessed Dementia Rating Scale. Marshall Folstein et al (1975) added praxis to his procedure, The Mini-Mental Status Examination, which was, and still is, extensively used in medical settings and research. Pfeiffer (1975), using both community- and in-hospital-based patients, developed and standardized the Portable Mental Status Questionnaire. Subsequently, Mattis' (1976) Dementia Rating Scale, and Jacob's et al (1977), Cognitive Capacity Screening Examination appeared in the literature. Hersch (1979) made available an extended scale of dementia, which has been deemed superior to Mattis' test (Lau et al 1988).

The literature has also demonstrated the need for simple and reliable procedures that can be conveniently administered in the office or at the bedside. A report by Williamson et al (1964) found that only 13% of demented patients living at home were recognized as such by their family physician. An investigation by Knights and Folstein (1977) demonstrated that cognitive impairment in hospital patients was recognized in 37% by their physicians and in 55% by their nurses. In a neurological population, only 70% of cognitively impaired patients were so diagnosed by neurologists (DePaulo and Folstein 1978).

The MMSE has been described elsewhere (Folstein et al 1975). It contains 11 items and gives a total score out of 30 and the cutting score is normally below 24. The NCSE does not have a total score, but rather a profile. This profile has 11 scales, each representing one aspect of cognitive functioning. An attractive feature of all the scales is that there is a screening item, which is deemed to be the most difficult level of the scale. This is followed by a series of questions of graded severity designed to measure the severity of deficits. If the patient is successful with the screening, the "metric" questions are not asked. If the screening is failed, the "metric" questions are pursued. This allows the test to concentrate on areas of deficit, potentially saving time. A standardization is

Table 2

Organic mental disorders

Diagnosis	Neurobehavioral Cognitive Status Examination			Mini-Mental State Examination		
	n	true positive	false negative	n	true positive	false negative
Delirium	2	1	1	0	N/A	N/A
Dementia	3	3	0	3	2	1
Organic mood	2	1	1	2	0	2
Organic amnestic	2	2	0	1	1	0
Organic personality	2	2	0	0	N/A	N/A
Intoxication	1	1	0	1	0	1

Table 3

Mood disorders

Diagnosis	Neurobehavioral Cognitive Status Examination			Mini-Mental State Examination		
	n	true negative	false positive	n	true negative	false positive
Major depression without psychotic features	24	12	12	14	14	0
Major depression with psychotic features	8	3	5	4	3	1
Mania	7	3	4	4	4	0
Other	4	2	2	1	1	0

displayed on the scoring sheet such that a normal score on any scale is recorded in a grey area. Everything which is out of this area may indicate cognitive dysfunction. The NCSE takes anywhere from ten to 20 minutes to complete.

OBJECTIVES OF THE STUDY

The objective of the study was to establish the validity of the NCSE as a case-finding test for organic mental disorders in seriously ill psychiatric patients. In the calculations of sensitivity, specificity and predictive value, clinical diagnoses were utilized as a gold standard. Given the objectives of the study, this type of gold standard was regarded as the most appropriate one. Other possible gold standards, for example neuropsychological testing, would be appropriate for determining the accuracy of the NCSE as a measure of cognitive dysfunction, but not its sensitivity and specificity for the detection of clinically relevant organic mental disorders.

METHODS

Seventy-two patients were recruited, with the help of their attending psychiatrists, from a clinical teaching unit located at the Calgary General Hospital. The admission criteria were as follows: patients who were ill enough to be admitted to hospital; patients who were over 18-years of age and patients who presented either with marked impairment of functioning, delusions, hallucinations, grossly disorganized behaviour or an impaired memory. These inclusion criteria (in conjunction with the exclusion criteria listed below) were intended to ensure that the sample resembled the clinical population in whom a structured test of cognitive functioning might be applied to help identify patients with organic mental disorders.

Because language skills and mental retardation can bias the results of cognitive tests, only patients whose primary language was English were included in the study. Patients with a diagnosis of mental retardation were excluded. Additional exclusion criteria included those patients who presented with an anxiety disorder a somatoform disorder or who had a principal diagnosis of a personality disorder. As discussed above, these criteria were formulated to help

ensure that the sample resembled the clinical population in whom the NCSE would most likely be applied.

Acceptance into the study included the following steps: a review of individual patients by the principal investigator with the referring physicians; an examination, by the principal investigator, of the information available in the clinical records of each potential subject; a determination as to whether each patient met the inclusion and exclusion criteria for the study, and finally, a clinical examination of the patient, by the principal investigator, to establish the DSM-III diagnosis. According to the DSM-III-R framework, organic mental disorders were diagnosed in the following way (American Psychiatric Association 1987): 1. recognizing the presence of one of the organic mental syndromes (delirium, dementia, intoxication, organic amnestic syndrome, organic hallucinosis, organic personality syndrome, organic anxiety syndrome, intoxication and withdrawal and 2. demonstrating, by means of the history, physical examination or laboratory tests, the presence of a specific organic factor judged to be etiologically related to the abnormal mental state.

After providing informed consent, all patients were given the NCSE by an independent examiner, who was blind to the admission diagnosis as well as to the DSM-III-R diagnosis established by the senior investigator. As in previous studies, the test results were considered abnormal if any of the 11 scales were outside of the normal range determined by the originator of the test. The protocol was also designed to produce a clinically realistic testing situation for a test oriented towards case-finding: one in which the tester was unaware of the diagnosis. In the second half of the study, all patients were given an MMSE immediately after the NCSE.

For the purpose of evaluating test-retest reliability, a number of patients were re-examined within a week following their admission. Similarly, a total of 20 patients were given an additional NCSE on by a second examiner for the purpose of testing inter-rater reliability.

ANALYSIS

The validity of the instrument was determined using calculations of sensitivity, specificity and predictive value in

Table 4

Psychotic disorders

	Neurobehavioral Cognitive Status Examination			Mini-Mental State Examination		
Diagnosis	n	true negative	false positive	n	true negative	false positive
Schizophrenia	9	3	6	7	7	0
Delusional disorder	1	1	0	0	N/A	N/A
Psychotic disorder not elsewhere classified	3	1	2	0	N/A	N/A

relation to the clinically determined diagnosis. For a subset of the patients, the MMSE was also performed in order to provide a standard for comparison. Reliability was estimated using the Kappa coefficient as an index of agreement between the test-retest and inter-rater comparisons.

RESULTS

Table 1 gives the DSM-III-R diagnosis for the total sample. The patients were 52% female and 48% male, 52% were between 26 and 45 years old and 35% were between 46 and 65-years-old. The percentage of the sample younger than 25 years and older than 65 years was 5% and 3% respectively.

Tables 2 through 5 display the performance of the instrument broken down by diagnostic group. One requirement of a case-finding instrument is that it be reasonably sensitive. In the current context, there should be a high probability of an abnormal score in those individuals who have an organic disorder. The performance of the two tests in patients with organic disorders is displayed in Table 2. Of the 12 patients with organic disorders, 10 had abnormal profiles on the NCSE, resulting in an overall sensitivity of 83%. This compared favourably with the MMSE, which detected only three of seven patients tested with this instrument (sensitivity = 43%). These finding are consistent with those of other recent studies, which have found the sensitivity of the NCSE to be greater than that of the MMSE. The negative predictive value (the probability that a patient who has passed the test does not have an organic disorder) was 93% for the NCSE and 89% for the MMSE.

The estimate of sensitivity for the MMSE is low relative to some previously published estimates, but the value obtained in this study is consistent with that of another study measuring the sensitivity of the MMSE in severely ill psychiatric patients (Lamarre and Patten 1991).

Another requirement for case finding instruments is that they be reasonably specific. In other words, patients presenting with non-organic conditions should have a high probability of passing the test. For the mood disorder category, there were 24 cases of major depression without psychotic features. Of these, the NCSE produced 12 abnormal results (false-positives) and 12 true negative results, resulting in a specificity of only 50%. Twelve of these patients were also

given the MMSE, and of these, all scored within the normal range, indicating 100% specificity. For major depression with psychotic features, the findings were similar. In this group, the MMSE again exhibited greater specificity than the NCSE; three of four (75%) *versus* three of eight (37.5%), respectively. The false positive result for the MMSE, occurring in a patient with psychotic depression was the only false positive result in 33 of the patients given this test in the study. The results for patients with psychotic disorders and other DSM-III-R disorders are summarized in Tables 4 and 5. In these patient groups, sample size limitations make comparison of the two tests difficult, but, in general, the NCSE was much less specific than the MMSE. The difference was particularly pronounced in patients suffering from schizophrenia , where only three of nine patients passed the NCSE, *versus* seven of seven who passed the MMSE. In terms of over all specificity, the NCSE produced a rate of 47%, which was considerably less than the 97% specificity of the MMSE.

Typically, when very stringent cut-off points are used to interpret diagnostic tests, the resulting interpretations have high specificity but low sensitivity. When less stringent cut-off points are used, there may be an increase in sensitivity, but at the cost of specificity. These relationships are usually evaluated using receiver-operator curves by plotting sensitivity and specificity *versus* potential cut-off scores. A traditional receiver-operator curve cannot be generated for NCSE data since the test does not utilize a summary score (instead the test results are presented as eleven scale scores). However, some patients fail the test by poor performance on only one scale, while others fail on more than one. Thus, it is possible to perform a receiver-operator analysis by plotting sensitivity and specificity versus the number of scales in which the subject performed outside of the normal range. A plot of this type is presented in Figure 1. When results in which a subject fails no more than one scale (rather than none) is scored as negative, specificity does increase somewhat (from 47% to 78%), but it does so at a considerable cost to sensitivity (from 83% to 58%). Progressively increasing the number of lower-than-normal scale scores required to classify a patient's performance as abnormal, results in progressively increased specificity, but with progressively lower sensitivity. Hence, altering the way in which the NCSE is scored does not appear to offer a solution to the problems

Table 5

Other disorders

Diagnosis	Neurobehavioral Cognitive Status Examination			Mini-Mental State Examination		
	n	true negative	false positive	n	true negative	false positive
Alcohol dependence	1	0	1	0	N/A	N/A
Mixed substance dependence	1	1	0	1	1	0
Adjustment disorder	1	1	0	1	1	0
V-code	1	1	0	1	1	0

noted with the sensitivity and specificity of the instrument in this population.

The positive predictive value of the NCSE was only 24%, meaning that an inability to pass the test increased to 24% the estimated probability that a patient in the study population had an organic disorder. This low predictive value could add little to clinical judgement because it is a small change from the pre-test estimated probability of a patient having an organic disorder (prevalence), which was 17% in this sample. In distinction, because of its greater specificity, the MMSE had a positive predictive value of 85%. Positive predictive value of a test often increases if the test is used in a population with a higher pre-test probability of disease. Therefore, the NCSE may have a higher positive predictive value in (for example) elderly patients. Unfortunately, there were insufficient numbers of elderly subjects (n = 2 over the age of 65) to evaluate this possibility.

Since the clinical diagnosis itself may be regarded as an imperfect measure (as opposed to being a gold standard), it may be appropriate to gauge agreement between clinical diagnosis and NCSE results using a Kappa coefficient, rather than estimates of sensitivity and specificity. The Kappa was 0.15, indicating weak agreement.

The test-retest reliability for the NCSE (Kappa) was 0.69, which suggests good reliability, but the inter-rater reliability Kappa value was only 0.57, which may be interpreted as moderate inter-rater agreement. Since the test-retest reliability estimate was based on a repeat testing within a week, and since patients' clinical status may have changed between testings, the estimated reliability may be less than the actual reliability.

CONCLUSIONS

The NCSE demonstrated very good sensitivity for the detection of organic mental disorders in this study. Overall, the sensitivity was 83%, which was considerably higher than the MMSE, 43%. Unfortunately, the improved sensitivity apparently came at the expense of specificity. The specificity of the NCSE was only 47%, whereas the specificity of the MMSE was 97%. It is a general tendency of screening and case-finding instruments to gain sensitivity at the expense of

specificity. In fact, at least in part, the weakness of the NCSE (poor specificity) is probably due to its strength; because minor cognitive impairment is frequent in patients with major non-organic psychiatric disorders, the NCSE was probably sensitive enough to detect this, resulting in poor specificity for the detection of organic disorders. The poor specificity of the NCSE resulted in a poor positive predictive value for the test.

Of the many available bedside instruments for the evaluation of cognitive functioning, the NCSE is the most sophisticated. A major strength of the test is its use of screening and metric items. This allows the test to concentrate on areas of impairment and spend less time testing areas where cognitive functioning is intact.

Because of this feature, the NCSE is brief enough to allow a busy clinician to complete a fairly sophisticated assessment of cognitive functioning as a part of his or her bedside assessment of the patient. It is our impression that the NCSE obtains superior information about cognitive functioning than would most of the other bedside instruments. However, the high quality of information provided by the NCSE does not translate into efficient case-finding for organic disorders in psychiatric inpatients.

The disappointing performance of the NCSE as a case-finding instrument in this study may reflect inherent limitations in the use of cognitive assessment instruments for the detection of organic mental disorders. Because some organic disorders, such as organic delusional disorder and organic mood disorder are typically characterized by minor cognitive dysfunction, and severe non-organic disorders may be accompanied by appreciable cognitive dysfunction, the use of cognitive screening instruments to detect organic disorders may be an approach with inherent limitations. Furthermore, some of the sets of criteria for organic mental disorders (such as organic mood disorder and organic delusional disorder) in DSM-IIIR are fairly vague, requiring for example, a judgement that the "disturbance" was initiated and maintained by a specific organic factor. Diagnostic errors in judgement could occur with the application of such criteria, and such errors could produce an appearance of low accuracy for the test.

Table 6

Summary of validity and reliability results

Validity	Sensitivity	NCSE	83%
		MMSE	43%
	Specificity	NCSE	47%
		MMSE	97%
Reliability	Test-retest	Kappa = 0.79	
	Inter-rater	Kappa = 0.57	

Another inherent limitation to the use of a single test result for case-finding is that some organic mental disorders, particularly Delirium, are characterized by a fluctuating sensorium. Fluctuations in mental state may account for the false negatives" among delirious patients in this study.

The NCSE displayed good test-retest reliability, but not as good inter-rater reliability in our evaluation. A post-hoc analysis revealed that the problems with inter-rater reliability were produced by several pairs of tests in which minor differences in scoring, usually on a single scale, resulted in an overall failure on the test. With the NCSE, a single abnormal result on a scale results in a classification of failure on the test. It should be noted that each rater studied the manual accompanying the test, but there were no standardized training or practice sessions. Institution of such procedures could potentially improve the inter-rater reliability.

Overall, the NCSE is capable of providing much useful clinical information. It is also short enough to be of practical use in psychiatry, and can be performed by clinicians at the bedside. However, it cannot be used alone as a screening test or case-finding instrument for the detection of organic mental disorders, at least not in a general psychiatry inpatient setting. It may, however, be useful to clinicians in other clinical contexts, and could be a used as a supplement to their clinical mental status examination.

ACKNOWLEDGEMENT

REFERENCES

American Psychiatric Association (1987) Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised. The American Psychiatric Association.

Blessed G, Tomlinson BE, Roth M (1968) The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. Br J Psychiatry 114:797-811.

DePaulo JR, Folstein MF (1978) Psychiatric disturbances in neurological patients: detection, recognition and hospital course. Annals of Neurology 4:225-228.

Folstein MF, Folstein SE, McHugh PR (1975) Mini-Mental State Examination. J Psychiatr Res 12:189-198.

Hersch EL (1979) Development and application of the extended scale for dementia. J Am Ger Soc 26: 348-354.

Jacobs JW, Bernhard MR, Delgado A, Strain JJ (1977) Screening for organic mental syndrome in the medically ill. Ann Intern Med 86:40-46.

Kahn RL, Goldfarb AI, Pollack M, Peck A (1960) Brief objective measures for the determination of mental status in the aged. Am J Psychiatry 117:326-328.

Knights EB, Folstein MF (1977) Unsuspected emotional and cognitive disturbance in medical patients. Annals Intern Med 87:723-724.

Lamarre CJ, Patten SB (1991) Evaluation of the Modified Mini-Mental Status in a general psychiatric population. Can J Psychiatry 36: 506-511.

Lau C, Wands K, Merskey H, Boniferro M, Carriere L, Fox H, Hachinski VC, et al (1988) Sensitivity and specificity of extended scale for dementia. Arch Neurol 45: 849-852.

Mattis S (1976) Mental status examination for organic Mental syndromes in the elderly patient. In: Bellak L, Karasu TE, (eds). Geriatric Psychiatry. New York: Gruen & Stratton.

Northern California Behavioral Group, Inc. (1963, 1986, 1988) Manual for the Neurobehavioral Cognitive Status Examination. Fairfax CA: Northern California Behavioral Group, Inc.

Pfeiffer E (1975). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 23: 433-441.

Schwann LH, Van Dyke C, Kiernan RJ, Edward LM, Mueller J (1987). The Neurobehavioral Cognitive Status Examination: comparison with the Cognitive Capacity Screening Examination and the Mini-Mental State Examination in a neurosurgical population. Ann Intern Med 107: 486-491.

Williamson J, Stokoe JH, Gray S, et al (1964) Old people at home: their unreported needs. Lancet 1:1117-1120.