INTRODUCTION

Neuropsychological tests are widely used to evaluate cognitive function in patients with brain injuries and neurological disorders such as dementia. In patients with dementia, neuropsychological testing is typically used to evaluate the presence and the severity of dementia, assign differential diagnoses, and monitor the disease progression and the drug effects (1).

The Seoul Neuropsychological Screening Battery (SNSB) is a comprehensive neuropsychological test that assesses five cognitive domains: attention, memory, language, visuospatial function, and frontal/executive function (2). The SNSB is one of the most commonly used neuropsychological tests in Korea for assessing cognitive functioning in patients with stroke, head trauma, Parkinson’s disease, and dementia. Although the SNSB is valid and reliable, with norms based on the analysis of 447 healthy subjects (2), several limitations of this battery have been identified. First, it has been suggested that the length of time taken to administer the test may be excessive for some patients with dementia, especially those with compromised attention spans. Second, although it is possible to evaluate each cognitive domain individually, the battery does not provide a global cognitive function (GCF) score, which may be useful for longitudinal monitoring of the disease progression. Third, it may be difficult to quantify the decline of frontal/executive function in patients of interest across time using SNSB, because the subtests assessing frontal/executive function include some tests that are qualitative in nature.

The main purpose of this study is to develop a dementia version of SNSB (SNSB-D) that was modified from the original SNSB such that it can be more easily applicable to patients with dementia for both clinical and research purposes. The SNSB-D differs from the original SNSB in the following ways: it provides a GCF score that represents a sum of the five cognitive domains assessed; it was shortened to reduce overall testing time; and it includes a method for converting qualitative items into a quantitative score. In order to evaluate the validity and reliability of the SNSB-D for assessing patients with memory impairment and dementia, we tested the SNSB-D in patients with mild cognitive impairment (MCI), Alzheimer’s disease (AD), and a normal control group (NC).
Seoul, Korea from November 2005 to January 2007. All patients with AD met the criteria for probable AD proposed by the National Institute of Neurological and Communicative Diseases and Stroke and Alzheimer’s disease and Related Disorders Association (NINCDS-ADRDA) (3). The aMCI patients were diagnosed according to the criteria proposed by Peterson et al. (4): 1) subjective memory complaint as described by the patient and/or caregiver, 2) normal general cognitive function, as defined by a score of 24 or greater on the Korean version of Mini-Mental State Examination (MMSE), 3) ability to participate in normal activities of daily living (ADL), judged clinically and by an ADL scale, 4) objective memory decline below the 16th percentile on neuropsychological tests, and 5) non-conformance to clinical criteria for diagnosis of dementia.

All patients underwent a comprehensive evaluation consisting of a detailed medical history, neurological examinations, and a neuropsychological evaluation (SNSB). Additionally, laboratory tests were used to confirm that there were no secondary causes for dementia or cognitive impairment. Magnetic Resonance Imaging (MRI) was performed on all patients, and all patients with structural brain lesions or severe white matter ischemia (caps or band >10 mm and deep white lesion >25 mm) were excluded. Patients who were illiterate were also excluded, regardless of formal education status.

The normal control (NC) group consisted of 77 healthy spouses or caregivers of patients from the memory disorder clinic. All controls were screened for neurological and psychiatric illnesses, and those who were identified to have any of these illnesses were excluded from the study. All subjects in the NC group met the criteria for healthy controls proposed by Christensen et al. (5) and did not have dementia as defined by the score below 8 points on the Korean Dementia Screening Questionnaire (KDSQ) (6) as well as by an ADL score less than 8 on Seoul Instrumental Activities of Daily Living (S-IADL) (7).

The three groups included in this study did not differ significantly in age, education level, and gender. In terms of Clinical Dementia Rating (CDR) scores (8), all participants in the NC group had a CDR score of 0; all patients in the MCI group had a CDR score of 0.5; and of the participants in the AD group 35 had a CDR score of 0.5 (mild stage, 38%), 42 had a CDR score of 1 (mild to moderate stage, 45%), and 16 had a CDR score of 2 (moderate stage, 17%). Among patients with AD, the CDR groups did not differ in age, education level, and gender. We obtained informed consents from all the patients and controls, and this study was approved by the Institutional Review Board of Samsung Medical Center (2005-02-008).

### Construction of SNSB-D

Like the original SNSB, the SNSB-D consists of sub-domains assessing attention, language and related function, visuospatial function, memory, and frontal/executive function (Table 1). We devised a GCF score drawn from the results of the SNSB-D that includes a maximum of 300 points, which is the sum of each sub-domain with 17/300 (6%) from attention, 27/300 (9%) from language and related function, 36/300 (12%) from visuospatial function, 150/300 (50%) from memory, and 70/300 (23%) from frontal/executive function.

Table 1 summarizes the differences between the SNSB-D and

| Domains                    | Score | %  | Subtests                                                                 | Maximum points | Modification from original SNSB |
|---------------------------|-------|----|--------------------------------------------------------------------------|----------------|---------------------------------|
| Attention                 | 17    | 6  | Digit span forward                                                      | 9              | None                            |
|                           |       |    | Digit span backward                                                     | 8              |                                  |
| Language and related function | 27    | 9  | Short form of K-BNT (A form) Calculation (3 items each for addition, subtraction, multiplication, division) | 15             | Short form instead of the full 60 item version of K-BNT Excluded items: spontaneous speech, auditory comprehension, repetition, reading, writing, four components of Gerstmann syndrome, limb and buccofacial praxis |
| Visuospatial function     | 36    | 12 | RCFT copy                                                               | 36             | None                            |
| Memory                    | 150   | 50 | Orientation                                                              | 6              | Out of 10 orientation items, season, country, city, and floor were excluded |
|                           |       |    | SVLT free/delayed recalls                                               | 48             |                                  |
|                           |       |    | SVLT recognition                                                        | 12             |                                  |
|                           |       |    | RCFT immediate/delayed recalls                                          | 72             |                                  |
|                           |       |    | RCFT recognition                                                        | 12             |                                  |
| Frontal/Executive function| 70    | 23 | Motor impersistence                                                     | 3              | Excluded items: alternating hand movement, alternating square and triangle, category word generation (supermarket), phonemic word generation (ㄱ, ㅅ), Stroop test-word reading |
|                           |       |    | Contrasting program                                                     | 3              |                                  |
|                           |       |    | Go-no-go test                                                           | 3              |                                  |
|                           |       |    | Fist-edge-palm                                                         | 3              |                                  |
|                           |       |    | Luria loop                                                              | 3              |                                  |
|                           |       |    | Category word generation (animal)                                       | 20             |                                  |
|                           |       |    | Phonemic word generation (ㄱ, ㅅ)                                         | 15             |                                  |
| GCF score                 | 300   | 100|                                                                           |                |                                  |

K-BNT, Korean-Boston naming test; RCFT, Rey-complex figure test; SVLT, Seoul verbal learning test; GCF, global cognitive function.
the original SNSB, and describes how the SNSB-D was modified from the original SNSB. The attention domain score was derived from the sum of the raw scores for forward and backward digit span. Of the many items that assess language and related functions, only naming and calculation tests were included. Specifically, a shortened version (15 item version, Form A) of the Korean version of the Boston Naming Test (K-BNT) and 12 trials of written calculations (three trials each for addition, subtraction, multiplication, and division) were included. The copying test from the Rey Complex Figure Test (RCFT) was used to assess visuospatial function. Several domains of memory function were assessed, including orientation, verbal memory, and visual memory. Orientation was assessed using four measures assessing orientation to time (year, month, date, day of the week) and two items assessing orientation to place (“What is this place for?” and “What’s the name of this place?”). The verbal memory score was derived from the sum score of three recall trials, a delayed recall task, and a recognition task from the Seoul Verbal Learning Test (SVLT). The visual memory score was composed of the sum of the scores from the immediate recall, delayed recall, and recognition scores of the RCFT. The frontal/executive function subdomain was assessed using motor impersistence, contrasting program, go-no-go test, fist-edge-palm task, and the Luria loop task, each of which were rated on a 0 to 3 scale. Additionally, a word fluency test and Stroop color-reading test were administered. For the word fluency test, a category word generation task (animal) and a phonemic word generation task (“‘/g/” were given, with a maximum score of 20 for the category word generation task and 15 for the phonemic word generation task. Thus, if the number of appropriate words generated was greater than 20 or 15, respectively, the measure was scored as 20 and 15 points. Stroop test score was converted to a maximum score of 20 by dividing the number of correct responses of color reading test by 5 and dropping the digits after the decimal points. Because of the shortening of several tests from the original SNSB as described in the above, it takes about 40-50 min to administer the SNSB-D compared with 60-80 min to complete the original SNSB.

### Test-retest reliability

To assess test-retest reliability, the SNSB-D was readministered one to three months (mean=51 days) following the initial test in 10 patients with AD, three patients with MCI, and eight participants from the NC group.

### Statistical analysis

Data were analyzed using SPSS 13.0. ANOVA was used to compare age and education levels across groups, and a chi-square test was used to compare gender across groups. To verify the discriminant validity of the SNSB-D, GCF scores were compared among all groups. To examine discriminant validity based on severity of dementia, all subjects were classified into four groups (CDR 0, 0.5, 1, 2), and GCF scores were compared among these four groups. To examine the convergent validity of the SNSB-D, Pearson correlation coefficients were calculated for the SNSB-D and the MMSE. The reliability of the SNSB-D was evaluated by assessing internal consistency and test-retest reliability. Cronbach’s alpha was used to test for internal consistency, and Pearson correlation coefficients were used to test for test-retest reliability among each of the groups.

To evaluate the diagnostic utility of the SNSB-D, we examined the sensitivity and specificity of SNSB-D, using a receiver operating characteristics (ROC) curve. Discrimination ability of GCF score derived from the SNSB-D was compared with the results of MMSE using area under the curve (AUC) analyses.

## RESULTS

### Discriminant validity by group analysis

GCF scores differed significantly among the groups (NC, MCI and AD; P<0.001; Table 2). Tukey’s post hoc analyses were used

| Group (CDR) | NC (0) | NC (0.5) | MCI | AD (1) | AD (2) | P value |
|-------------|--------|----------|-----|--------|--------|---------|
| MMSE        | 28.6±1.4 | 27.2±1.9 | (20.2±3.7) | 19.2±5.0 (19.0±4.0) | (13.4±4.6) | <0.001 |
| Attention   | 10.0±2.1 | 9.4±2.1  | (8.3±2.1) | 7.6±2.1 (7.4±2.0) | (6.3±2.0) | <0.001 |
| Language and related function | 24.1±2.8 | 21.1±3.7  | (18.1±5.2) | 15.5±5.7 (15.6±4.5) | (9.7±5.5) | <0.001 |
| Visuospatial function | 33.8±2.5 | 32.2±4.0  | (26.8±11.1) | 21.5±12.3 (22.0±10.5) | (8.4±9.7) | <0.001 |
| Memory      | 88.3±15.3 | 56.0±16.3 | (31.8±11.3) | 24.6±11.9 (23.0±10.2) | (13.0±5.0) | <0.001 |
| Verbal memory | 39.6±7.2 | 27.6±8.2  | (16.6±5.2) | 13.2±6.2 (12.7±5.6) | (7.3±4.7) | <0.001 |
| Visual memory | 42.7±12.4 | 22.8±12.9 | (10.9±7.5) | 8.2±6.9 (7.6±6.6) | (3.8±3.4) | <0.001 |
| Frontal/Executive function | 54.8±9.2 | 48.2±10.9 | (35.8±10.6) | 28.1±12.2 (25.6±10.1) | (17.9±10.7) | <0.001 |
| GCF score   | 211.0±24.3 | 166.9±28.9 | (120.7±28.3) | 97.2±35.5 (93.6±27.1) | (55.2±26.6) | <0.001 |

NC, normal control; MCI, mild cognitive impairment; AD, Alzheimer’s disease; MMSE, Mini-Mental State Examination; SNSB-D, dementia version of Seoul Neuropsychological Screening Battery; GCF, global cognitive function.

DOI: 10.3346/jkms.2010.25.7.1071  
http://jkms.org  
1073
to examine differences in GCF scores between each of the groups. GCF scores were significantly lower for the AD group than for the MCI group \((P<0.001)\), and GCF scores were significantly lower for the MCI group than for the NC group \((P<0.001)\). In contrast, a post hoc analysis of MMSE scores revealed that there were no statistically significant differences between the NC and MCI groups \((P=0.105)\), and that the only significant difference in MMSE scores was between the MCI and AD groups \((P<0.001)\).

With respect to each of the cognitive domain scores, the three groups differed in each of the domains assessed: attention, language and related function, visuospatial function, memory, and frontal/executive function \((all \, P<0.001)\). Tukey’s post hoc analyses of the domain scores revealed that there were significant differences between the NC and MCI groups, as well as between the MCI and AD groups in language and related function, memory, and frontal/executive function. However, in attention and visuospatial function, differences were only noted between the MCI and AD groups, but not between the NC and MCI groups.

The same analyses were performed to examine each of the sub-test scores individually. Overall group differences were present in all tests at the significance level of \(P<0.001\), with the exception of the motor impersistence test \((P=0.072)\) from the frontal/executive function sub-domain. Post hoc analyses revealed that there were significant differences between the NC and MCI groups, as well as between the MCI and AD groups on the K-BNT; the immediate recall, delayed recall, and the recognition of the SVLT; the immediate and delayed recall, as well as the recognition of the RCFT; and the category word generation test. However, in the remaining sub-tests, including forward digit span, backward digit span, the calculation task, the RCFT copy and orientation task, contrasting program, go-no-go test, fist-edge-palm test, the Luria loop task, and the word fluency test, there were statistically significant differences between the MCI and AD groups, but not between the NC and MCI groups.

**Discriminant validity by CDR stage analysis**

Participants were also classified according to the CDR score. As the CDR stage increased, the GCF score decreased: CDR 0 \((211.0\pm24.3)\), 0.5 \((146.1\pm36.7)\), 1 \((93.6\pm27.1)\), 2 \((55.2\pm26.6)\), and GCF scores differed significantly between the four CDR groups. There were also significant differences between each of the consecutive pairs of CDR groups: CDR 0 vs. 0.5, CDR 0.5 vs. 1, and CDR 1 vs. 2 \((P<0.001)\).

To further examine the utility of the SNSB-D to discriminate according to dementia severity rating, participants were reclassified into five groups according to CDR scores. Specifically, the NC group was assigned a CDR score of 0 (NC-0), MCI patients were assigned a CDR score of 0.5 (MCI-0.5), and AD patients were assigned a CDR score of 0.5 (AD-0.5), 1 (AD-1), or 2 (AD-2), depending on the patient’s severity of dementia. According to post hoc analyses, the five CDR groups differed significantly with respect to GCF scores. Specifically, the GCF score decreased significantly as the CDR increased (Table 2). We were particularly interested in comparing the MCI-0.5 and AD-0.5 groups.

Post hoc analyses revealed that the two groups differed significantly across all cognitive domains except attention.

**Convergent validity**

In examining convergent validity, it was determined that SNSB-D GCF scores and MMSE scores were highly correlated \((r=0.876)\). Additionally, each of the cognitive sub-domain scores of the SNSB-D was significantly correlated with the overall GCF score, as well as with the MMSE score. Specifically, it was found that the memory domain score had the highest correlation with the
GCF among the five cognitive sub-domains, with the other correlations as follows: attention (0.629), language and related function (0.848), visuospatial function (0.779), memory (0.945), frontal/executive function (0.919).

**Reliability**

Cronbach's alphas for the SNSB-D were 0.797 in the AD group, 0.710 in the MCI group, 0.637 in the NC group, and 0.871 for the full sample, suggesting a high level internal consistency. The test-retest reliability of the GCF score was assessed using Pearson correlation coefficients. Test-retest reliability was high, with a correlation coefficient of 0.978 for the whole sample, 0.918 for the AD group, 0.999 for the MCI group, and 0.960 for the NC group.

**Diagnostic utility**

We used the ROC curve analysis and calculated the area under the curve (AUC), as a measure of diagnostic utility for the SNSB-D. First, in discriminating the MCI group from the NC group, the AUC of the SNSB-D was 0.883 and the SNSB-D had a sensitivity of 0.844 and a specificity of 0.814, with a cut-off score of 188.25 (Table 3, Fig. 1A). The AUC of the SNSB-D GCF score was significantly superior to the MMSE (AUC=0.716, P=0.001). Second, in discriminating the MCI group from the AD group, the AUC of the SNSB-D was 0.939 and the appropriate SNSB-D cut-off score was 134.25 with a sensitivity of 0.860 and a specificity of 0.860 (Table 3, Fig. 1B). The AUC of the MMSE (0.942) was comparable to that of the SNSB-D (P=0.842).

**DISCUSSION**

The SNSB-D is a neuropsychological test battery designed to assess attention, language and related function, visuospatial function, memory, and frontal/executive function. Summing across the cognitive sub-domains results in a global cognitive functioning score. For constructing a GCF score from the cognitive sub-domain scores, the ideal proportion of each domain score is not known. Therefore, we thought it was worthwhile to compare the construct and proportion of cognitive domain scores of SNSB-D with those of other tests that have global and subscale scores and the results are as follows. As shown in Table 4, the SNSB-D was compared with the MMSE (9), Mattis Dementia Rating Scale (DRS) (10), Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-cog) (11), and Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) (12). All prior tests, with the exception of the DRS, share the feature of having a memory domain which represents approximately 50% of the total score. The proportions of the other cognitive domains comprising the global score were variable among tests. The MMSE does not have tests for frontal/executive function (9). The DRS contains no tests for language function, and compared to memory function it assigns a higher proportion to executive function (10). The ADAS-cog does not include items that measure attention or frontal/executive function (11). Recently a global scoring system for the CERAD was proposed by Chandler et al. (12), however, this measure does not have an attention sub-domain. In comparison to these tests, the SNSB-D contains tests that assess attention (6%), memory (50%), language (9%), visuospatial function (12%), and frontal-executive function (23%). Given that memory decline is the earliest change observed in most patients with dementia and MCI, the proportion of 50% in the memory domain is likely to be appropriate.

The SNSB-D was shown to have good internal consistency and high test-retest reliability. Additionally, we wanted to see if the SNSB-D GCF score could discriminate among the NC, MCI

---

**Table 3. AUCs and cut-off scores for the SNSB-D GCF score and the MMSE global score**

| Global score | NC and MCI | MCI and AD |
|--------------|------------|------------|
| **SNSB-D global score** | | |
| AUC | 0.883 | 0.939 |
| Cut-off | 188.25 | 134.25 |
| Sensitivity | 0.844 | 0.860 |
| Specificity | 0.814 | 0.860 |
| **MMSE global score** | | |
| AUC | 0.716 | 0.942 |
| Cut-off | 28.5 | 25.5 |
| Sensitivity | 0.610 | 0.837 |
| Specificity | 0.698 | 0.882 |

AUC, area under the curve; MMSE, Mini-Mental State Examination; SNSB-D, dementia version of Seoul Neuropsychological Screening Battery; GCF, global cognitive function.

**Table 4. Cognitive sub-domain composition of the neuropsychological tests**

| Cognitive function | MMSE | DRS | ADAS-cog | CERAD | SNSB-D |
|--------------------|------|-----|----------|-------|--------|
| Points (%) | Points (%) | Points (%) | Points (%) | Points (%) | Points (%) |
| Attention | 5 (17) | 37 (26) | - | - | - | 17 (6) |
| Language | 8 (27) | - | 25 (36) | 15 (15) | 17 (9) |
| Visuospatial function | 1 (3) | 6 (4) | 5 (7) | 11 (11) | 36 (12) |
| Praxis | - | - | 5 (7) | - | - |
| Memory & orientation | 16 (53) | 25 (17) | 35 (50) | 50 (50) | 150 (50) |
| Frontal/Executive function | - | 76 (53) | - | 24 (24) | 70 (23) |
| Global scores (%) | 30 (100) | 144 (100) | 70 (100) | 100 (100) | 300 (100) |

MMSE, Mini-Mental State Examination; DRS, Dementia Rating Scale; ADAS-cog, Alzheimer’s Disease Assessment Scale-Cognitive Subscale; CERAD, Consortium to Establish a Registry for Alzheimer’s Disease; SNSB-D, dementia version of the Seoul Neuropsychological Screening Battery.

DOI: 10.3346/jkms.2010.25.7.1071
and AD groups. We found that the GCF was able to distinguish each group effectively. In contrast, the MMSE score was unable to discriminate between the MCI (27.2±1.9) and NC (28.6±1.4) groups; although it was able to discriminate between the MCI and AD groups. GCF scores could also discriminate between each of the CDR stages. GCF scores declined significantly as CDR scores increased, confirming that the SNSB-D is able to discriminate the stages of the CDR. Both MCI and early stages of AD can be assigned a CDR score of 0.5. Thus, we were particularly interested in whether the GCF score might differentiate between MCI patients and AD patients with CDR scores of 0.5. The results showed that these two groups differed significantly in terms of the GCF scores. Therefore, the SNSB-D may serve as a useful tool that can discriminate not only MCI, AD, and normal aging, but also distinguish AD patients at different stages on the CDR scale.

We considered the diagnostic utility of the GCF scores derived from the SNSB-D by comparing the SNSB-D with the MMSE, using a ROC curve analysis. Generally, an AUC above 0.85 is regarded as a strong diagnostic value (13). In discriminating between the NC and MCI groups, the AUC of the SNSB-D was 0.883, whereas the AUC of the MMSE was 0.716, suggesting that the utility of SNSB-D is superior to that of MMSE for differentiating MCI from controls. Likewise, the SNSB-D had AUC values above 0.85 when discriminating among patients with MCI and AD at differing severity levels, indicating that the SNSB-D is more sensitive than the MMSE in discriminating between MCI and the early stages of AD.

Although the SNSB has been widely used to diagnose dementia for clinical, as well as research purposes in Korea, the absence of a global score has resulted in a lack of longitudinal, follow-up investigations of cognitive function in patients with dementia. In this study, we have demonstrated that the SNSB-D, with its global and domain-specific scores, is a valid and reliable instrument for assessing individuals ranging from normal levels of cognitive function to MCI or AD. This new instrument will also be helpful to monitor the progression of global and domain-specific cognition over time. Furthermore, although the utility of SNSB-D has not been verified in other dementing disorders, since the SNSB-D consists of tests that cover various cognitive domains, the GCF score and sub-domain scores of SNSB-D could be used to assess and monitor cognitive dysfunctions in patients with other types of dementia.

REFERENCES

1. Lezak MD. Neuropsychological assessment, 3rd ed. New York: Oxford University Press 1995.
2. Kang Y, Na DL. Seoul Neuropsychological Screening Battery (SNSB). Seoul: Human Brain Research & Consulting Co; 2003.
3. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer’s disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer’s Disease. Neurology 1984; 34: 939-44.
4. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999; 56: 303-8.
5. Christensen KJ, Multhaup KS, Nordstrom S, Voss K. A cognitive battery for dementia: development and measurement characteristics. J Consult Clin Psychol 1993; 3: 168-74.
6. Yang DW, Cho BL, Chey JY, Kim SY, Kim BS. The development and validation of Korean dementia Screening questionnaire (KDSQ). J Korean Neurol Assoc 2002; 20: 135-41.
7. Ku HM, Kim JH, Kwon EJ, Kim SH, Lee HS, Ko HJ, Jo S, Kim DK. A study on the reliability and validity of Seoul-Instrumental Activities of Daily Living(S-IADL). J Korean Neuropsychiatr Assoc 2004; 43: 189-99.
8. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1999; 43: 2412-4.
9. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975; 12: 189-98.
10. Mattis S. Dementia Rating Scale (DRS). Odessa, FL: Psychological Assessment Resources 1988.
11. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer’s disease. Am J Psychiatry 1984; 141: 1356-64.
12. Chandler MJ, Lacritz LH, Hynan LS, Barnard HD, Allen G, Deschner M, Weiner MF, Cullum CM. A total score for the CERAD neuropsychological battery. Neurology 2005; 65: 102-6.
13. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978; 8: 283-98.