Journal of NeuroVirology, 11: 503–511, 2005

ISSN: 1355-0284 print / 1538-2443 online
DOI: 10.1080/13550280500384966

Taylor & Francis
Taylor & Francis Group

Clinical validation of the NeuroScreen

Ronald J Ellis,[1] Scott R Evans,[2] David B Clifford,[3] Lauren R Moo,[4] Justin C McArthur,[4] Ann C Collier,[5] Constance Benson,[6] Ron Bosch,[7] David Simpson,[8] Constantin T Yiannoutsos,[9] Yijun Yang,[2] Kevin Robertson[10] for the Neurological AIDS Research Consortium and the AIDS Clinical Trials Group Study Teams A5001 and A362

Departments of [1]Neurosciences and [6]Medicine, UCSD AntiViral Research Center (AVRC) and HIV Neurobehavioral Research Center, University of California, San Diego, California, USA; [2]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA; [3]Department of Neurology, Washington University, St. Louis, Missouri, USA; [4]Departments of Neurology and Epidemiology, Johns Hopkins University, Baltimore, Maryland, USA; [5]Department of Medicine, University of Washington School of Medicine, Seattle, Washington, USA; [7]Statistical and Data Analysis Center, Harvard School of Public Health, Boston, Massachusetts, USA; [8]Department of Neurology, Mount Sinai Medical Center, New York, New York, USA; [9]Division of Biostatistics, Indiana University School of Medicine, Indianapolis, Indiana, USA; [10]Department of Neurology, School of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA

The NeuroScreen comprises two easily administered components: the Brief NeuroCognitive Screen (BNCS), designed to estimate the frequency of human immunodeficiency virus (HIV)-associated cognitive disorders; and the Brief Peripheral Neuropathy Screen (BPNS), for distal sensory polyneuropathy (DSPN) in HIV. In this study, both the NeuroScreen and a more extensive standardized validation neurodiagnostic evaluation were administered to HIV-positive subjects ($N = 301$) enrolled in two large cohort studies at multiple sites. BNCS performance was summarized in the form of a demographically adjusted mean *z*-score, the NPZ3. The area under the receiver-operating characteristic (ROC) curve for the BNCS as compared to the reference standard neuropsychological (NP) evaluation was 0.74 (95% confidence interval [CI] 0.69, 0.79). Using a cut-point of −0.33 on the NPZ3 provided a correct classification rate of 68%, with roughly balanced sensitivity (65%) and specificity (72%). Under the assumption of a 30% prevalence of cognitive impairment, the calculated positive predictive value (PPV) of the BNCS was 86%. Relative to its reference standard, a modified Total Neuropathy Score (TNS) administered by a neurologist, the BPNS gave a similar correct diagnostic classification rate of 78%; sensitivity 49% [95% CI 37%, 60%]; specificity 88% [95% CI 82%, 91%]. Under the assumption of a 40% prevalence of DSPN, the PPV of the BPNS was 72%. These predictive values suggest that the NeuroScreen will be useful for tracking trends in the prevalence of HIV-associated neurologic disease in large cohorts in the era of combination antiretroviral therapy. However, because it yields substantial numbers of false positives and negatives, the NeuroScreen may be less useful in evaluating individual patients. *Journal of NeuroVirology* (2005) **11**, 503–511.

Keywords: antiretroviral; cognitive impairment; HIV; neuropathy

Address correspondence to Ronald J. Ellis, Department of Neurosciences, UCSD AntiViral Research Center and HIV Neurobehavioral Research Center, University of California, San Diego, 150 W. Washington Street, San Diego, CA 92103, USA. E-mail: roellis@ucsd.edu

This work was supported by AI38858 (AIDS Clinical Trials Group [ACTG]), and U01 NS0322228 to the Neurologic AIDS Research Consortium (NARC), Statistical and Data Management Center grant U01 AI038855 to the Harvard University Center for Biostatistics in AIDS Research, and NS049465 (to JCM).

The authors gratefully acknowledge the ACTG 362 and A5001 teams for referring subjects, the assistance of Mariana Cherner, PhD, and the UCSD HIV Neurobehavioral Research Center in conducting the training of site personnel for this study. The following sites accrued subjects to this study: Johns Hopkins University; University of California, San Diego; Mount Sinai Medical Center, New York; University of Rochester; Los Angeles County/University of Southern California Medical Center; University of Washington, Seattle; Washington University, St. Louis; Northwestern University; Cook County Hospital; University of North Carolina; University of Texas Southwestern Medical Center; University of Hawaii at Manoa, Leahi hospital; University of Colorado Health Sciences Center; University of Pennsylvania, Philadelphia; Cornell University.

Received 7 March 2005; revised 28 May 2005; accepted 31 May 2005.

Introduction

Numerous studies indicate that combination antiretroviral therapy has substantially impacted neurological disease in human immunodeficiency virus (HIV) infection (Cysique *et al*, 2004; Dore *et al*, 1999; Maschke *et al*, 2000; Morgello *et al*, 2004). Severe dementia has become less frequent, with milder forms of cognitive impairment assuming greater importance (McArthur, 2004). Mild neurocognitive disorders often occur despite successful systemic immune reconstitution. Sensory polyneuropathy also remains a common problem, and is etiologically often related to the use of neurotoxic nucleosides (Morgello *et al*, 2004). Thus although neurological morbidity has been altered substantially by highly active antiretroviral therapy (HAART), neurological disorders remain common and important in HIV infection. There is a need for brief screening instruments that adequately assess neurological function. The purpose of this study was to evaluate the performance (i.e., predictive values) of such a brief tool for neurological screening.

The NeuroScreen was designed to identify neurological disorders with relatively high prevalence and impact in the setting of HIV infection. It is an inexpensive and brief tool intended to provide rapid, reliable results and to be applicable for use by nonneurological personnel in large numbers of patients at multiple sites. The overall objective of this study was to validate use of the NeuroScreen administered by nonphysicians to ascertain trends in neurological disease incidence and prevalence in the HAART era, and to delineate strengths and weaknesses of the NeuroScreen Instrument. We evaluated the clinical diagnostic utility of the NeuroScreen by comparison to a more comprehensive neurodiagnostic evaluation, including neurologist assessment. The BNCS was validated by comparison to norm-based neuropsychological (NP) impairment classifications and clinical ratings. The BPNS was validated by comparison to a modified Total Neuropathy Score (TNS), as performed by a neurologist experienced in the assessment of HIV-infected patients.

Results

The study enrolled 301 subjects between August 2001 and September 2002. The median number of subjects enrolled at each of the 15 sites was 24 (interquartile range [IQR] 8, 28; range 1, 45). Demographically, the study sample was mostly non-Hispanic white (60%) men (86%), with a mean age of 44 years (standard deviation [SD] = 9). African Americans comprised 26% of the subjects, and other ethnic minorities 14%. The mean educational level was 14 years (SD = 3); 31% of subjects had a high school education or less and 37% had completed college or higher. A past history of intravenous (IV) drug use was reported by 13% of subjects. The CD4 nadir (cells/mm^3; median [IQR]) was 105 [28, 273]; and current median CD4 was 417 [280, 631].

Assessment of premorbid intellectual functioning
Scores on the WAIS-3 Vocabulary test, an indicator of educational attainment and premorbid intellectual functioning, were unimpaired (not more than 1 SD below demographically adjusted expectations) in 243 subjects (81%). Only 13 subjects (4%) had Vocabulary scores more than 2 SD below expected.

Rates of impairment according to the validation battery
Using an objective, performance-based criterion (at least 1 SD on two tests or 2 SD on one test), the overall rate of NP impairment on the validation battery was 56% (170/301 subjects). Using clinical ratings, which by design adjust for limitations in the normative data and potential adverse conditions of test administration, the rate of NP impairment was somewhat lower (76 subjects; 25%).

NeuroScreen—Cognitive Portion (BNCS)
Impairment cut-points: When screening neurocognitive performance (BNCS) was rated according to a conventional impairment criterion of at least 1 SD below the mean on each of two tests, or 2 SD on one test, 43 of the 301 subjects (14.9%) were classified as impaired. Sensitivity and specificity compared to the validation battery (at least one SD on two tests or 2 SD on one test) were 23.6% (95% CI 17.8%, 30.7%) and 98.3% (95% CI 94.0%, 99.5%). Under the assumption of a prevalence of impairment in the underlying population equal to that seen in the study sample, positive predictive and negative predictive values were 95.1% and 47.9%. Using a less stringent impairment classification criterion of at least 1 SD on one test, 94 subjects (31.8%) were impaired; sensitivity was 44% and specificity 84%, with a correct classification rate of 60%. Figure 1 illustrates sensitivity and specificity according to these cut-points and compares them to NPZ3 cut-points as plotted by the receiver-operating characteristics (ROC) method (see below).

Averaged z-scores (NPZ3): The median NPZ3 in the subject sample was −0.32 (IQR: −0.86, 0.23). To evaluate the performance of the screening tests relative to the reference standard impairment classification from the validation battery, an ROC curve was constructed. This curve plots performance characteristics (sensitivity, specificity) for varying screening NPZ3 cut-points. Figure 1 compares the NPZ3 ROC curve to one derived using the simple, impairment points method described above. The area under the NPZ3 ROC curve (AUC) was 0.74 (95% CI 0.69, 0.79), significantly greater than that calculated for the impairment point method (0.654; 95% CI 0.60, 0.71; $P = .0003$).

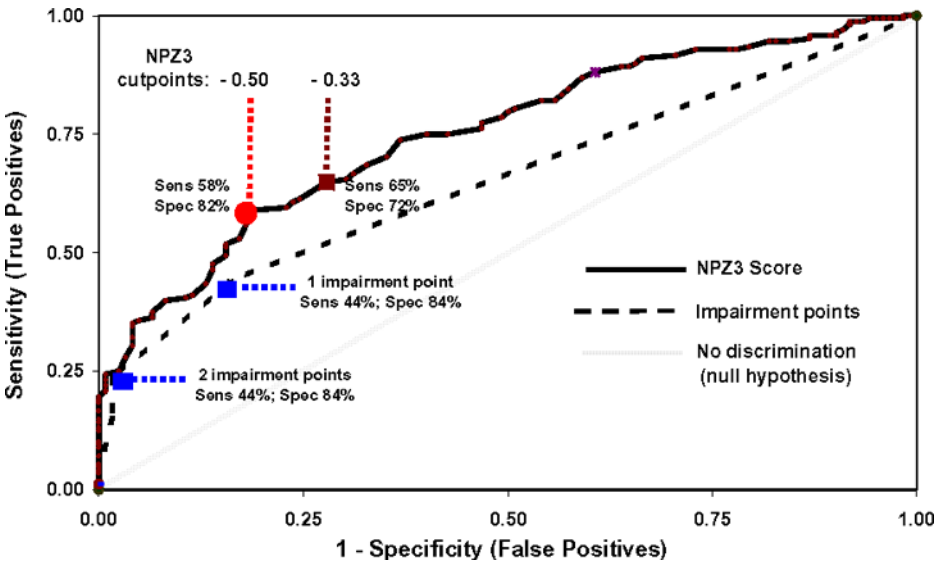

Figure 1 ROC curves for the cognitive portion of the NeuroScreen using NPZ3 scores and impairment points (IP) as compared to a reference standard (norm-based neurocognitive impairment classification). For the NPZ3 ROC, the area under the curve was 0.74 (95% CI 0.69, 79), whereas for the impairment points ROC, the area under the curve was 0.65.

Table 1 lists the sensitivity and specificity for several possible cut points for the impairment point and NPZ3 methods. The NPZ3 cut-point that provided the highest overall correct classification rate was −0.5 (sensitivity 58%; specificity 82%; 68% correct classification). A NPZ3 cut-point of −0.33 also provided a correct classification rate of 68% (sensitivity 65%; specificity 72%), and a cut-point of 0.0 generated a correct classification rate of 67% (sensitivity 78%; specificity 51%).

We also evaluated a revised, "deficit score" NPZ3 method in which positive *z*-scores (above-average performance) for any of the three BNCS tests were reassigned a value of 0 when computing the NPZ3.

Table 1 Sensitivity and specificity of the BNCS relative to the Validation Examination, using impairment cut-points and averaged *z*-scores (NPZ3)

Cut-point	*% sensitivity*	*% specificity*	*% correctly classified*
Impairment point method			
≥0	100.0	0.0	58.5
≥1	44.0	83.9	60.6
≥2	23.5	98.3	54.6
≥3	15.1	98.3	49.7
≥4	4.2	98.3	44.0
≥5	1.8	100.0	42.6
≥6	1.2	100.0	42.3
NPZ3 method			
≤3.0	100.0	0.0	57.9
≤2.0	100.0	0.8	58.3
≤1.0	96.4	9.8	60.0
≤0.0	78.6	50.8	66.9
≤−0.5	58.3	82.0	68.3
≤−1.0	35.1	95.9	60.7
≤−1.5	14.9	100.0	50.7
≤−2.0	4.8	100.0	44.8
≤−2.5	0.6	100.0	42.4

The pattern of findings when using this method did not differ substantially from the conventional NPZ3 score, and results are not reported here.

Although applying the neuropsychologist's clinical rating method as the gold standard yielded a lower prevalence of impairment (25%) as previously noted, performance characteristics of the screening tests were similar to those derived using norm-based classification. Using a cutpoint NPZ3 score of 0.5, the overall correct classification rate was 68%; sensitivity 68%; specificity 68%.

Prevalence of syndromic HIV-related neurocognitive disorders: The neurocognitive tests in the validation battery, although designed to be sensitive to HIV-associated neurocognitive disorders, will also detect impairment related to comorbid conditions such as head injury and developmental learning disabilities. Moreover, the tests may detect impairment in individuals who do not recognize any neurocognitive difficulties. Syndromic neurocognitive disorders (HIV-associated minor cognitive-motor disorder [MCMD] and HIV-associated dementia [HAD]) are those in which patients are aware of some cognitive difficulties, and an experienced clinician judges that the etiologic contribution of HIV itself is at least as important as other contributing causes. To enhance the likelihood that different clinicians would apply diagnostic criteria in a similar fashion, we employed an algorithmic approach. Table 2 summarizes the prevalence of HIV-associated MCMD and HAD among the 170 subjects classified as impaired by the Validation Battery. Among 43 subjects classified as impaired with the BNCS, 7 cases were diagnosed as having MCMD, 2 cases HAD, and the remaining 34 cases had NP impairment, but did not meet criteria for a syndromic diagnosis.

Table 2 Prevalence of syndromic neurocognitive disorders in the 170 subjects impaired on the NP validation battery

	N	%
NP impairment, does not meet criteria for syndromic disorder	82	48.2
Possible or probable MCMD	27	15.9
Possible or probable HAD	4	2.4
Probable NP impairment due to other cause	14	8.2
Other, unable to reliably assign diagnosis	43	25.3
Total	170	

MCMD, HIV-associated minor cognitive-motor disorder; HAD, HIV-associated dementia. The remaining 131 unimpaired subjects were classified as neurocognitively normal.

Neuropathy (BPNS)

BPNS screening data were available for 292 (97%) of the 301 subjects. Of these, 77 (26%) had diminished vibratory sensation in the toes bilaterally, and 98 (34%) had absent or hypoactive ankle reflexes bilaterally. Sixty subjects (21%) met objective criteria for "suspected DSPN," comprising both distal vibratory loss and diminished ankle reflexes bilaterally.

TNS data were available for all 301 subjects. Of these, 68 (23%) met criteria for neuropathy. Ten subjects (3%) had symptoms suggestive of neuropathy, but no objective findings supporting the diagnosis, and four subjects (1%) had only one neuropathic sign, i.e., insufficient to meet diagnostic criteria. Relative to the TNS as a standard, the BPNS gave an overall correct diagnostic classification rate of 78%; sensitivity 49% (95% CI 37%, 60%); specificity 88% (95% CI 82%, 91%). Assuming a prevalence of neuropathy in the population equal to that in the study sample, positive predictive and negative predictive values were 53% and 85%. The low sensitivity but high specificity of the BPNS suggested that the criteria used to gauge the BPNS were too stringent, and may have excluded true positive cases. To address this possibility, we investigated the effect on the performance characteristics of revising the definition of neuropathy from "at least mild loss of vibration sensation in both great toes bilaterally ***and*** ankle reflexes absent or hypoactive relative to knees bilaterally" to "at least mild loss of vibration sensation in both great toes bilaterally ***or*** ankle reflexes absent or hypoactive relative to knees bilaterally." Using this definition, sensitivity improved to 80%; however, as expected, specificity decreased to 59%. Also in an attempt to identify the source of disagreement between the screening and validation neuropathy exams, we compared the evaluation of ankle-reflex loss from the nurses' (BPNS) and the neurologists' (TNS) assessments. The sensitivity and specificity of the nurses' ankle reflex assessments were 52% and 73% respectively.

Table 3 tabulates key findings from the neurologists' clinical examinations. The neurologists found motor signs and symptoms to be relatively uncommon. Moderate weakness was reported by 5% of subjects and slight difficulty by 12%. Neurologists found moderate weakness in only 1% of subjects and mild weakness in 6%.

Table 3 Key findings from the neurologists' clinical examinations, according to the Total Neuropathy Score method

	N	%
Sensory symptoms (paresthesias, numbness, neuropathic pain)		
None	184	61%
Limited to the tips of fingers or toes	68	23%
Symptoms extend to ankle or wrist	26	9%
Symptoms extend from above ankle or wrist to the level of knee or elbow	19	6%
Symptoms above knee or elbow and/or severe disabling symptoms	4	1%
Sensory function (distal vibration)		
Normal	128	42
Absent/decreased in fingers or toes	71	24
Absent/decreased up to wrist or ankle	67	22
Absent/decreased up to elbow or knee	34	11
Absent/decreased above the level of elbow or knee	1	0
Ankle reflexes		
Normal/brisk	140	47
Pathologically increased	11	4
Reduced	86	29
Absent	64	21

Clinicians' etiologic diagnostic classifications: The peripheral nerve evaluation used in this study (BPNS and TNS) was designed to detect neuropathic signs and symptoms related to any disease condition. In addition to HIV, other etiologies, such as diabetes mellitus, alcohol abuse, and nutritional deficiencies, might contribute to the presence of peripheral nerve dysfunction. To address etiologic diagnoses, an experienced clinician formulated a diagnostic interpretation in which the contribution of HIV was weighed against other possible contributing factors in each patient. To enhance the likelihood that different clinicians would apply diagnostic criteria in a similar fashion, we employed an algorithmic approach. A diagnosis of HIV-associated distal sensory polyneuropathy (DSPN) indicated that HIV itself was believed to be the most important ("probable DSPN") or at least as important (possible "DSPN") as other contributing causes. A diagnosis of nucleoside-associated DSPN indicated that the clinician judged that neuropathy symptoms onset was related to use of neurotoxic nucleoside reverse transcriptase inhibitors (d4T, ddI, ddC) and that dose reduction or discontinuation of the offending agent provided at least partial symptomatic relief. Of the 60 subjects whose physical findings indicated DSPN, information necessary to formulate a clinical etiologic diagnosis was available for 57. Table 4 shows the etiologic classifications of these subjects.

Discussion

The NeuroScreen was designed to optimize cost, resource efficiency, simplicity, rapid availability of

Table 4 Clinical diagnostic classification of 57 subjects meeting criteria for DSPN on the TNS

Diagnosis	*N*	*%*
Possible or probable HIV-associated DSPN	22	39
Nucleoside-associated neuropathy	22	39
DSPN, other cause	2	4
Unable to confidently assign an etiologic diagnosis	11	19
Total	57	

results, adaptability to a variety of clinical settings, and acceptability in the population being studied. In the present study, we evaluated the NeuroScreen's performance characteristics relative to a reference standard diagnostic evaluation, specifically, a validation assessment consisting of a more comprehensive neurological and NP diagnostic evaluation performed by neurologists and neuropsychologists experienced in the assessment of HIV-infected subjects.

The NeuroScreen, as applied in this study, was designed to estimate population prevalence rather than to formulate individual diagnoses. In light of this intended use, we reasoned that balanced performance characteristics (i.e., roughly equal sensitivity and specificity) were most appropriate. Using the NPZ3 method to classify neurocognitive screening performance in a sample of 301 neurologically unselected, HIV-infected individuals, we found that an NPZ3 cut-point of −0.33 provided a reasonable balance of sensitivity (65%) and specificity (72%), and a correct classification rate of 68%. Similarly, the neuropathy screen gave a sensitivity of 49%, specificity 88% and an overall correct diagnostic classification rate of 78%.

As effective combination antiretroviral therapy has transformed HIV into a chronic, manageable disease, the relative importance of neurologic morbidity has increased (Dore *et al*, 2003; Sacktor *et al*, 2002). The performance characteristics of the NeuroScreen will make it useful for examining trends in the prevalence and incidence of disease in large cohorts, and for assessing factors that might modify disease occurrence, such as antiretroviral therapy effectiveness or toxicity in the central nervous system (CNS). Additionally, the NeuroScreen may be used to assist in selection of special populations for more intensive study or for treatment trials. For example, screening of participants at multiple sites could be used to identify a subset of individuals potentially eligible for a new neuroprotective treatment.

The documented performance characteristics of the NeuroScreen, although adequate for the intended use of tracking population prevalence, are not appropriate for the detection of a very serious disease in individual patients. In such a situation, false-negative results (failure to detect disease when it is present) can be extremely costly.

A variety of modifications and enhancements might be proposed to achieve the goal of improving the performance characteristics of the NeuroScreen. For example, the specific NP tests might be broadened to include other domains of cognitive ability sensitive to HIV-related brain dysfunction, such as learning and executive abilities. For neuropathy, more sophisticated tests of peripheral nerve function such as electrophysiological studies might provide better performance characteristics. Additionally, performance characteristics might improve if the tests were administered by more highly trained personnel, such as expert neurologists and neuropsychologists. However, such modifications would increase the cost of the assessments and might make them less widely applicable.

The performance characteristics of the NeuroScreen should be evaluated in the context of available alternative screening instruments. Carey *et al* (2004) found somewhat better neurocognitive screening performance characteristics (sensitivity = 78% and 75%) using the Hopkins Verbal Learning Test—Revised (HVLT-R; Total Recall), a test of learning and memory, combined with either the Grooved Pegboard Test nondominant hand (PND), a test of motor function, or the WAIS-III Digit Symbol (DS), a test of mental flexibility very similar to the WAIS-R Digit Symbol used in the present study. These measures were administered to 190 HIV-positive participants at a single site, different from the large, multisite cohort we evaluated.

The cognitive cutoffs (BNCS NPZ3) that achieved conventionally acceptable levels of sensitivity and specificity were surprisingly lenient. Thus, a cutoff NPZ3 of −0.33, reflecting an average deficit of only one third of the standard deviation below normative performance, yielded the best overall impairment classification rate (68%), with nearly balanced sensitivity and specificity (65% and 72%). This finding suggests that the specific screening tests chosen are "easy" relative to those that detected impairment in the validation battery. Also, this result is consistent with milder overall levels of impairment in this group of predominantly HAART-experienced subjects.

We evaluated two alternatives to the NPZ3 method for classifying subjects as impaired or unimpaired. The impairment point method is simple and easy to use, but demonstrated low sensitivity (24%). Moreover, although the impairment point method might be convenient for practitioners, it is known to suffer from downward bias (Hanley and McNeil, 1982). The NPZ3 method provided better sensitivity, although sacrificing specificity only to a moderate degree. Neither deficit scores nor clinical ratings substantially altered diagnostic efficiency.

Another limitation of the NeuroScreen is that it identifies only the presence or absence of abnormalities, not their cause. Impairment on NP testing may be caused by confounding disorders other than HIV infection, such as drug intoxication, learning disability, prior heavy substance abuse, or chronic hepatitis C infection. Because conditions such as these are

quite common among individuals with HIV infection, it was not practical or desirable to exclude all such subjects from an analysis whose goal was to evaluate utility of the screening tests in large populations. Thus, individuals showing impairment on the NeuroScreen would still require a diagnostic evaluation to determine the etiology of their disorder and to select appropriate treatment.

Some subjects in this study had been tested using the NeuroScreen one or more times before entering this study. Prior experience with NP screening tests is associated with improved performance, usually referred to as "practice effects." Performance improvement due to practice would make a subject less likely to be classified as impaired, resulting in relatively conservative prevalence estimates for neurocognitive impairment.

Although increasing numbers of HIV-infected individuals are surviving with neurologic disorders, access to expert and neurological and NP evaluations is often limited, even in research settings. Thus the need for effective screening tests is pressing. The Adult AIDS Clinical Trials Group (AACTG) NeuroScreen is a brief, simple, low-cost screening tool for HIV neurologic disease designed to screen for prevalent or incident cases of HIV-related neurologic impairment in populations at risk. Because of its demonstrated simplicity and adaptability to multisite administration, it may be particularly useful in resource-limited settings where appropriate normative comparison data are available. The classification error rates derived from this study will allow adjusted and unbiased estimates of disease prevalence to be obtained, improving upon "naïve" estimates based on observed frequencies of cognitive impairment and DSPN. Thus the NeuroScreen can now be used to more reliably estimate disease prevalence and potentially to track changes over time, yielding vital information about the impact of HAART on the burden of neurological disease in HIV.

Materials and methods

Design
This multicenter, cross-sectional study was conducted at 15 NeuroAIDS Research Consortium (NARC) sites with expertise in performing NP and neurological evaluations in individuals with HIV infection. To reduce referral bias due to overrepresentation of subjects from individual sites, enrollment at each site was limited to 60 subjects. Each subject underwent two separate evaluations conducted by different personnel: a screening assessment and a validation assessment.

Subjects
Individuals eligible to participate in this validation study were HIV-infected volunteers enrolled in each of two qualifying AIDS Clinical Trials Group (ACTG) protocols as described below. Exclusion criteria were (1) active major psychiatric or medical illness likely to interfere with the subjects' ability to complete the study; (2) active substance abuse or other factors that could prevent compliance, at the discretion of the local investigator. ACTG site personnel informed subjects of their potential eligibility for participation in NARC007, and if a subject granted permission to be referred, NARC personnel obtained consent. Enrollment in this study occurred within three months after completion of the BNCS and the BPNS in one of the qualifying study visits.

Qualifying protocols
The qualifying protocols were two ACTG longitudinal studies selected to sample individuals with a wide range of HIV disease stages and antiretroviral interventions. Subjects in the first, "A5001: Adult AIDS Clinical Trials Group Longitudinal Linked Randomized Trials (ALLRT) Protocol," participated in a variety of randomized clinical intervention trials, including trials of combination antiretroviral therapy. The second qualifying protocol was "A362: A Trial to Determine Time to Development of AIDS-Defining Complications and Serious Infections in Subjects Who Have Achieved an Increase in CD4 cells on Antiretroviral Therapy." Subjects in this study had documented histories of severe immunocompromise and had undergone immune reconstitution on HAART (CD4 count nadir <50 and recovered to >100) before entry.

Training and testing materials
Prior to formal training, site examiners were provided with standardized test source documents and case report forms, written instructions, and videotapes illustrating test administration. The study organizers sponsored an in-person NeuroScreen training session, which consisted of a review of the test procedures, followed by a hands-on workshop in which instructors observed examiners administering the tests and provided feedback on their performance. Examinees were required to answer all questions carefully on a 20-question, multiple-choice self-assessment test. The training materials were available for training of newly hired personnel. Detailed validation study procedures were presented to neurologists and neuropsychologists from each of the participating sites. A standard set of normative tables was used for this study.

Overview of clinical evaluations
The screening tests used in this study were designed to be relatively inexpensive, resource efficient, simple, and reliable, and to yield quick results and be practical for administration to large numbers of patients at diverse locations. By comparison, the validation tests represented a robust standard (collectively termed the Validation Assessment), that included a

neurological examination administered by a neurologist and NP testing supervised by a clinical neuropsychologist. The Validation Assessment consisted of two parts: the Neurocognitive Diagnostic Evaluation to assess HIV cognitive-motor disorders (HIV-CMD) and the Neuropathy Diagnostic Evaluation to assess DSPN.

Screening evaluations
The NeuroScreen consisted of a series of short tests designed to elicit findings indicative of diagnoses of HIV-cognitive motor disease (the BNCS) and DSPN (BPNS) in HIV-infected individuals. Administration of the neurocognitive screen required 12 to 15 min, and the neuropathy screen took 2 to 3 min. The cognitive portion of the NeuroScreen consisted of the Trailmaking test (parts A and B), and the WAIS-R Digit Symbol task, which together assess speed of information processing, mental flexibility, and working memory.

Overall performance on the BNCS was summarized by two methods: impairment points (IP) and NPZ3. IP were assigned by comparing subjects' raw scores norms from an HIV-negative reference sample, adjusted for age, gender, and sex (Heaton, 1992; Heaton *et al*, 1991). Raw scores (RS) were converted to IP as follows: RS 1 to 2 SD below norms = 1 IP; RS > 2 SD below norms = 2 IP; otherwise = 0 IP. The impairment sum was the sum of impairment points for each of the three tests. Overall performance on the BNCS was defined as "impaired" when the impairment sum was 2 or higher. The NPZ3 was calculated as the mean of the *z*-scores on the three BNCS tests. A *z*-score for each test was calculated by subtracting a normative mean (as described above) from each subject's score and then dividing by the appropriate standard deviation for each test (also estimated from the normative sample). Norms published by Psychological Assessment Resources in February 2004 (Heaton *et al*, 2004) were utilized and adjusted for age, education, sex, and ethnicity.

The neuropathy portion of the NeuroScreen comprised evaluation of distal vibratory sensation loss and diminished ankle reflexes. Nurse coordinators at each site were trained in the assessment of distal vibration and ankle reflexes. Subjects were classified as having DSPN if they had at least mild loss of vibratory sensation in both great toes bilaterally, and ankle reflexes absent or hypoactive relative to knees bilaterally. Details of a similar assessment tool have been published previously (Marra *et al*, 1998).

Validation assessments
NP validation battery: The tests comprising the validation NP battery and the principal neurocognitive domains assessed by these tests are listed in Table 5. A test of premorbid ability (WAIS-III Vocabulary) was included to assist in determining whether subjects' impaired performance was affected by a limited educational experience or other factors unrelated to HIV that were not reflected in the demographic characteristics. There was no overlap between tests included in the screening and validation NP batteries.

Table 5 Tests comprising the validation neuropsychological (NP) examination

Test	*Domain assessed*
WAIS-3 Vocabulary	Premorbid ability
Timed gait	Motor skills
Grooved Pegboard—dominant and nondominant	Motor skills
HVLT-R Learning Trials	Verbal learning
HVLT-R Delay	Memory (delayed recall)
HVLT-R Recognition	Memory (recognition)
WAIS-3 Symbol Search	Speed of information processing
Paced Auditory Serial Addition Test—1 channel	Attention/working memory

WAIS-3 = Wechsler Adult Intelligence Scale, 3rd edition.
HVLT-R = Hopkins Verbal Learning Test—Revised, Form A.

NP impairment classification
To assess the utility of the neurocognitive portion of the NeuroScreen, we compared it to two reference standard impairment classification systems: norm-based classification and clinical ratings. Statistical analyses examined the performance characteristics of the screening tests against both impairment classification systems in parallel. By norm-based classification, the reference standard for impairment was performance at least 1 SD below demographic norms on two or more tests or at least 2 SD below norms on at one test. This approach has the advantage of being objective and unbiased. However, clinicians frequently accommodate relative deficiencies in available normative data by applying a clinical rating system for impairment classification. For this study, clinical ratings (Heaton, 1994) were performed by an experienced senior neuropsychologist (KR). The neuropsychologist was provided with each subject's demographic information, raw test scores, T-scores, and behavioral notes, but remained blinded to information on treatment status, disease stage and other identifying information. The clinician classified each participant's global NP status on the following 9-point scale: 1 = above average functioning; 2 = average; 3 = below average; 4 = borderline/atypical; 5 = definite mild impairment; 6 = mild to moderate impairment; 7 = moderate impairment; 8 = moderate to severe impairment; and 9 = severe impairment. A global rating of 5 or above, indicating overall abnormal NP functioning, required a rating of 5 or higher in at least two ability areas (i.e., a single, isolated ability deficit did not qualify for a global rating within the "impaired" range).

In addition, a consensus diagnostic process was used to formulate diagnostic classifications based on the neurocognitive validation evaluation using

criteria described by the American Academy of Neurology task force (ANN, 1991). Subjects whose performance was less than 1 SD below norms on all tests, or between 1 SD and 2 SD below norms on only one test, were designated NP normal. HAD was diagnosed when a subject performed more than 1.5 SD below norms on at least two tests and showed substantial impairment in activities of daily living. MCMD was diagnosed when a subject showed impairment of at least 1 SD on two tests or 2 SD on one test and demonstrated at least some impairment of instrumental activities of daily living. Subjects showing impaired test performance but reporting no difficulty in activities of daily living were designated "NP impaired" (NPI), without further specification. Subjects meeting criteria for NP impairment but in whom the impairment was judged likely to be caused by a process other than HIV, were not eligible to receive diagnoses of HAD, MCMD, or NPI, and were classified separately.

Neuropathy validation evaluation

The neuropathy validation assessment tool, a modified TNS (Cornblath *et al*, 1999), comprised a clinical evaluation performed by a neurologist and recorded on standardized clinical record forms. The TNS is simple, inexpensive, brief, and yet has been shown sensitive to the development of distal neuropathy and to have excellent intra- and interrater reliability (Cornblath *et al*, 1999). The modified TNS used in this study omitted the nerve conduction studies and quantitative sensory testing. Two symptom domains (sensory, motor) and three objective examination domains (sensory, motor, and reflexes) were assessed. The symptom questionnaire took approximately 3 min to complete, and the directed examination took approximately 5 min. Symptoms and objective examination findings were combined to yield an overall index of neuropathy severity. A diagnosis of neuropathy was made if a subject obtained an abnormal score (>1 point) in three or more domains.

Statistical analyses

Validation of the screening instruments (BNCS, BPNS) was carried out by comparing their respective disease classifications (impaired versus not impaired; neuropathy versus no neuropathy) to the corresponding reference standard classifications. Measures of evaluation included sensitivity, specificity, false-positive and false-negative rates, and predictive values with corresponding 95% CIs. CIs for sensitivity and specificity were produced with the Wilson score method without continuity correction (Newcombe, 1998). CIs for positive and negative predictive values were calculated with the method described by Simel and colleagues (Simel *et al*, 1991).

ROC curves plot the sensitivity versus (1-specificity) and illustrate the relationships between sensitivity and specificity as screening test cutoffs were varied. ROC curves were compared and CIs for the area under the curve were calculated using methods of DeLong *et al* (1998), which incorporates the within-subject correlations between the impairment points and NPZ3 scores.

Significance testing was performed at the .05 level with no adjustment for multiple testing and thus results should be interpreted with caution. All reported P values are for two-sided tests.

References

ANN (1991). Nomenclature and research case definitions for neurologic manifestations of human immunodeficiency virus-type 1 (HIV-1) infection. Report of a Working Group of the American Academy of Neurology AIDS Task Force. *Neurology* **41:** 778–785.

Carey CL, Woods SP, Rippeth JD, Gonzalez R, Moore DJ, Marcotte TD, Grant I, Heaton RK (2004). Initial validation of a screening battery for the detection of HIV-associated cognitive impairment. *Clin Neuropsychol* **18:** 234–248.

Cornblath DR, Chaudhry V, Carter K, Lee D, Seysedadr M, Miernicki M, Joh T (1999). Total neuropathy score: validation and reliability study. *Neurology* **53:** 1660–1664.

Cysique LA, Maruff P, Brew BJ (2004). Antiretroviral therapy in HIV infection: are neurologically active drugs important? *Arch Neurol* **61:** 1699–1704.

DeLong ER, DeLong DM, Clarke-Pearson DL (1998). Comparing the areas under two or more correlated receiver operating curves: a nonparametric approach. *Biometrics* **44:** 837–845.

Dore GJ, Correll PK, Li, Y, Kaldor JM, Cooper DA, Brew BJ (1999). Changes to AIDS dementia complex in the era of highly active antiretroviral therapy. *AIDS* **13:** 1249–1253.

Dore GJ, McDonald A, Li, Y, Kaldor JM, Brew BJ (2003). Marked improvement in survival following AIDS dementia complex in the era of highly active antiretroviral therapy. *AIDS* **17:** 1539–1545.

Hanley JA, McNeil BJ (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143:** 29–36.

Heaton RK (1992). Comprehensive norms for an expanded Halstead-Reitan battery: a supplement for the WAIS-R. In: *Psychological assessment resources*. Odessa, FL.

Heaton RK (1994). The utility of clinical ratings for detecting early cognitive change in HIV infection. In: *Neuropsychology of HIV infection*. Grant I, Martin A (eds). New York: Oxford University Press: pp 188–206.

Heaton RK, Grant I, Matthews CG (1991). Comprehensive norms for an expanded Halstead-Reitan battery: demographic corrections, research findings, and clinical applications. In: *Psychological assessment resources*. Odessa, FL.

Heaton RK, Miller SW, Taylor MJ, Grant I (2004). Revised comprehensive norms for an expanded Halstead-Reitan battery: demographically adjusted neuropsychological Norms for African American and Caucasian adults. In: *Psychological assessment resources*. Lutz, FL.

Marra CM, Boutin P, Collier AC (1998). Screening for distal sensory peripheral neuropathy in HIV-infected persons in research and clinical settings. *Neurology* **51:** 1678–1681.

Maschke M, Kastrup O, Esser S, Ross B, Hengge U, Hufnagel A (2000). Incidence and prevalence of neurological disorders associated with HIV since the introduction of highly active antiretroviral therapy (HAART). *J Neurol Neurosurg Psychiatry* **69:** 376–380.

McArthur, JC (2004). HIV dementia: an evolving disease. *J Neuroimmunol* **157:** 3–10.

Morgello S, Estanislao L, Simpson D, Geraci A, DiRocco A, Gerits P, Ryan E, Yakoushina T, Khan S, Mahboob R, Naseer M, Dorfman D, Sharp V (2004). HIV-associated distal sensory polyneuropathy in the era of highly active antiretroviral therapy: the Manhattan HIV Brain Bank. *Arch Neurol* **61:** 546–551.

Newcombe RG (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* **17:** 857–872.

Sacktor N, McDermott MP, Marder K, Schifitto G, Selnes OA, McArthur JC, Stern Y, Albert S, Palumbo D, Kieburtz K, De Marcaida JA, Cohen B, Epstein L (2002). HIV-associated cognitive impairment before and after the advent of combination therapy. *J NeuroVirol* **8:** 136–142.

Simel DL, Samsa GP, Matchar DB (1991). Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* **44:** 763–770.