NeuroTrax™ Science Letters
Volume 2, Issue 3; October 2004

Message from the Editor:

In this issue, the critical question of what constitutes normal performance on Mindstreams tests is examined. Research data collected from patients with an array of cognitive diagnoses was analyzed, and a set of performance ranges was defined based on the likelihood of abnormality. The resulting scale is important as it represents progress in the application of cognitive neurology to a broad range of diseases.

Ely Simon, MD

In this issue:

Defining 'Normal' Performance on Mindstreams Tests: A Comprehensive Analysis of NeuroTrax Research Data

GM Doniger, PhD[1], A Schweiger, PhD[2], ES Simon, MD[1], and the Neurocognitive Study Group*

ABSTRACT

Objective: To utilize data collected in controlled research studies over a range of cognitively abnormal diagnoses to determine an appropriate normal/abnormal cutoff for Mindstreams summary measures.

Background: In keeping with conventional neuropsychological practice, the 'normal range' of performance on Mindstreams computerized cognitive tests was initially set at ±1 standard deviation (SD) about an age- and education-appropriate normative mean. However, the suitability of this range has not been examined. Indeed the choice of an appropriate normal/abnormal cutoff is influenced by such factors as congruity between the normative sample and the study sample, currentness of the normative sample, and the relative severity of false positives (FP) and false negatives (FN). The present analysis was designed to identify an appropriate normal/abnormal cutoff for detecting mild impairment with Mindstreams tests on the basis of research data from individuals with a wide array of cognitive diagnoses and an appropriate normative sample.

Methods: Participants were 822 individuals enrolled in controlled research studies using Mindstreams™ (NeuroTrax Corp., NY) tests. Individuals received an expert diagnosis of cognitively healthy or a diagnosis indicative of specific forms of cognitive impairment. Mindstreams 'index scores' summarizing performance in particular cognitive domains (e.g., memory, executive function) were the primary dependent variable. Index scores predicted to evidence impairment for each cognitive diagnosis were identified and the performance of cognitively healthy individuals compared with that of cognitively impaired individuals, for each diagnosis individually ('individual cell comparisons') and across identified diagnoses ('cross-cell comparisons'). Mindstreams measures were normalized according to age and education and fit to an IQ-style scale (mean: 100; SD: 15), with the normative sample drawn from the same research databases as the study sample. Between-groups differences were tested and FP and FN rate (p[FP], p[FN]) computed at six normal/abnormal cutoffs relative to the normative mean. For each comparison, cutoffs meeting each of three criteria were identified: a) p(FP) closest to 0.10, b) p(FN) closest to 0.10, and c) $|p(\text{FP})-p(\text{FN})|$ smallest. The cutoff meeting each criterion across the majority of comparisons was identified, separately for individual and cross-cell comparisons.

Results: For the majority of individual and cross-cell comparisons, robust ($p<0.001$) between-group differences were found, criterion *a* was met at -1SD, criterion *b* at +0.25SD, and criterion *c* at -0.25SD.

Conclusions: -0.25 SD units was adopted as a best-balance normal/abnormal cutoff, with equivalent severity of FP and FN. To improve clinical utility, -1SD was adopted to distinguish between abnormal and 'probable abnormal' and +0.25SD was adopted to differentiate between 'probable normal' and normal. The resulting set of performance sub-ranges constitutes a powerful clinical tool for gauging probability of cognitive impairment associated with a wide array of cognitive diagnoses on the basis of Mindstreams testing.

[1] Dept. of Clinical Science, NeuroTrax Corporation, New York, NY USA
[2] Dept. of Behavioral Sciences, Academic College of Tel Aviv, Tel Aviv, Israel

What is the 'normal range' of performance on Mindstreams computerized cognitive tests? In keeping with conventional neuropsychological practice, the normal range was initially set at ±1 standard deviation (SD) about the age- and education-appropriate normative mean. A score of -1 to -2SD is generally considered borderline abnormal on traditional neuropsychological tests, including such tests as the Rey Auditory Verbal Learning Test (RAVLT) and the California Verbal Learning Test (CVLT) often used to detect mild cognitive impairment (MCI; Ratcliff & Saxton, 1998; Spreen & Strauss, 1998; Helmes, 2000). Indeed Petersen's criterion for objective impairment in MCI-amnestic has been operationalized as -1 or -1.5SD on a neuropsychological test of memory (Petersen et al., 1999; Palmer et al., 2003; Ganguli et al., 2004; Grundman et al., 2004).

Despite the wide acceptance of -1 to -2SD as normal/abnormal cutoffs, it is clear that the suitability of a given cutoff is directly related to the composition of the normative sample. The normative data may either be too broad or too restricted for comparison to the clinical population of interest. Most often the normative sample is community-based, including individuals with an array of demographic profiles, and is thus too broad. Compounding their over-inclusiveness, community-based samples sometimes use self-rating and hence may include individuals with undiagnosed pathology. Alternatively, the normative sample may be too restrictive. As Spreen and Struass (1998) point out, "patients are often tested in a hospital setting with attendant anxiety about ongoing illness and with less than optimal attention and motivation; dependent on the type of hospital, such patients may also be far from representative of the general population." In the case of a commonly used set of CVLT norms (Delis et al., 1987), the authors caution that their highly educated sample may be too restrictive. For the same normalized score (e.g., -1SD), actual performance will be higher or lower than appropriate if the normative sample is an inappropriate reference for the clinical population. In the case of the Delis et al. (1987) CVLT norms, -1SD is likely associated with a raw score that is actually average, but comes out abnormal because the normative sample is highly educated. Another factor contributing to incongruity between the normative sample and the clinical population of interest is their datedness. Indeed traditional neuropsychological tests are often scored against normative data collected decades earlier and therefore no longer representative of current normal individuals.

In addition to the appropriateness of the normative sample, the suitability of a given normal/abnormal cutoff is tied to the base rate and the relative severity of false positives and false negatives (or the relative importance of specificity versus sensitivity). As Ratcliff and Saxton (1998) point out, "less stringent criteria seem to be appropriate in older individuals, in whom impairment is more common, or when the purpose of the evaluation is to screen individuals prior to possible referral for more elaborate workup." Hence the conventional cutoff of -1 to -1.5SD in MCI may actually be influenced by a relatively high base rate (3-20% depending upon criteria; Busse et al., 2003) and low severity of false negatives given that the objective deficit is only one of a set of criteria for MCI and will likely be accompanied by an extensive workup to confirm/disconfirm the classification made by the neuropsychological test alone. To date there are no studies of the false positive and false negative rates of neuropsychological tests in MCI detection at the conventional cutoffs or otherwise, but the need for such studies is recognized (Rivas-Vasquez et al., 2004).

Given these considerations, the present analysis sought to define an appropriate 'normal range' for Mindstreams tests. Good congruity was achieved between the study sample and the normative sample in that both were drawn from databases of the same controlled research studies. Hence, the normative sample was not overly broad or restrictive, nor was the normative data outdated relative to the study data. The study sample consisted of patients with a variety of diagnoses associated with cognitive impairment and cognitively healthy individuals. In analyzing a range of disease-specific cognitive deficits, three normal/abnormal cutoffs were identified, each with a different relative severity of false positives and false negatives. An approach was developed utilizing these multiple cutoffs to define not only a 'normal range', but a set of performance ranges that enhance the clinical utility of Mindstreams tests.

Methods

Analyses were conducted on data from 822 participants in controlled research studies (including those of the Neurocognitive Study Group*) using Mindstreams™ (NeuroTrax Corp., NY) computerized tests (Dwolatzky et al., 2003). Each participant received an expert diagnosis (Table 1), which was taken as the gold standard. Expert diagnoses were based on the judgment of physicians relying on

Table 1. Distribution of expert diagnoses for participants in the study sample (*N*=822).

Designation	Diagnosis	*N*
Normal	Cognitively Healthy	401
Abnormal	MCI	128
	ADHD	98
	TBI	74
	Mild Dementia	41
	Parkinson's Disease	31
	Schizophrenia	16
	HLGD	15
	Dyslexia	6
	Other	12

ADHD: Attention Deficit Hyperactivity Disorder
HLGD: High Level Gait Disorder
MCI: Mild Cognitive Impairment
TBI: Traumatic Brain Injury

patient history, physical examination, and ancillary laboratory or imaging data, as necessary. For patients with multiple visits, only data from the first visit was included. Only patients whose primary language (i.e., most comfortable using, language used most often) was available as a Mindstreams test language were included.

The NeuroTrax system has been described elsewhere (Dwolatzky et al., 2003). In brief, Mindstreams consists of custom software that resides on the local testing computer and serves as a platform for interactive cognitive tests that produce accuracy and reaction time (RT; millisecond timescale) data. Once tests are run on the local computer, data are automatically uploaded to a central sever, where calculation of outcome parameters from raw single-trial data and report generation occur.

The tests (Table 2) sample various cognitive domains, including memory (verbal and non-verbal), executive function, visual spatial skills, verbal fluency, attention, information processing, and motor skills. All responses were made with the mouse or with the number pad on the keyboard. Patients were familiarized with these input devices at the beginning of the battery, and practice sessions prior to the individual tests instructed them regarding the particular responses required for each test.

Outcome parameters varied with each test. Given the speed-accuracy tradeoff (Cauraugh, 1990), a performance index (computed as [accuracy/RT]*100) was computed for timed Mindstreams tests in an

Table 2. Mindstreams tests for detection of mild impairment.

Go-NoGo Response Inhibition
Expanded Go-NoGo Response Inhibition
Verbal Memory
Non-Verbal Memory
Problem Solving
Stroop Interference
Finger Tapping
Catch Game
Staged Information Processing Speed
Verbal Function
Visual Spatial Orientation

attempt to capture performance both in terms of accuracy and RT. To minimize differences in age and education and to permit averaging performance across different types of outcome parameters (e.g., accuracy, RT), each NeuroTrax outcome parameter was normalized and fit to an IQ-style scale (mean: 100, SD: 15) in an age- and education-specific fashion (see Normalization below).

Index Scores

Normalized subsets of outcome parameters were averaged to produce seven summary scores as follows, each indexing a different cognitive domain:

MEMORY: mean accuracies for learning and delayed recognition phases of Verbal and Non-Verbal Memory tests

EXECUTIVE FUNCTION: performance indices (accuracy divided by RT) for Stroop Interference test and Go-NoGo Response Inhibition (either standard or expanded) test, mean weighted accuracy for Catch Game

VISUAL-SPATIAL: mean accuracy for Visual Spatial Orientation test

VERBAL: weighted accuracy for verbal rhyming test (part of Verbal Function test)

ATTENTION: mean reaction times for Go-NoGo Response Inhibition (either standard or expanded) and choice reaction time (a non-interference phase of the Stroop test) tests, mean reaction time for a low-load stage of Staged Information Processing Speed test, mean accuracy for a medium-load stage of Staged Information Processing Speed test

Table 3. Difference in mean performance between cognitively healthy (*N*=401) and patients with various cognitive diagnoses for Mindstreams index scores. Differences are in normalized units (SD: 15) computed according to age and education.

	Difference from Normal							
Mindstreams™ Measure	MCI	ADHD	TBI	Mild Dementia	Parkinson's Disease	Schizophrenia	HLGD	Dyslexia
Memory	12.74	8.96	16.43	20.56	7.86	25.65	9.51	-0.05
Executive Function	8.04	7.90	13.59	14.51	7.98	23.78	9.12	9.85
Visual Spatial	9.03	6.34	5.77	10.75	21.52	15.27	12.21	-1.49
Verbal Function	12.96	4.69	20.61	19.81	-6.71	14.69	NA	9.47
Attention	6.78	10.55	14.17	15.57	6.38	29.22	4.65	5.59
Information Processing Speed	8.21	11.75	14.49	16.01	7.77	23.10	NA	11.38
Motor Skills	4.23	2.00	11.00	9.34	8.75	25.44	5.72	6.16

SD: standard deviation
NA: Not Tested or *N*<6
Shading indicates comparisons to be used in further analyses for which cognitive impairment was anticipated in the abnormal group.

INFORMATION PROCESSING SPEED: performance indices (accuracy divided by RT) for various low- and medium-load stages of the Staged Information Processing Speed test

MOTOR SKILLS: mean time until first move for Catch Game, mean right and left inter-tap intervals for Finger Tapping test

These seven index scores served as the primary dependent variables for the present analysis. A Global Cognitive Score (GCS) computed as the average of these index scores served as a secondary dependent measure.

As batteries differed in the tests administered, data for all outcome parameters was not present for all patients. Missing outcome parameter data was also attributable to invalidation by quality control mechanisms triggered by response patterns indicative of poor compliance with test instructions (e.g., too many trials with the same response). Memory, Executive Function, Attention, and Motor Skills index scores were computed only if data was present for at least two of their constituent outcome parameters. The Information Processing Speed index score was computed only if data was present for at least three of its constituent outcome parameters, and the GCS was only computed only if data was present for at least three index scores.

Analysis

For each index score, the difference in mean performance was computed between cognitively healthy (*N*=401) and each individual abnormal diagnosis (Table 3). Cognitive domains predicted to evidence a performance decrement for specific diagnoses were identified on the basis of prior studies (Table 3, shaded cells), and analyses restricted solely to these cells. Analyses were conducted both for each cell individually ('individual cell comparisons') and across all shaded cells for a given index score ('cross-cell comparisons').

As an example, for the Memory index score, only the following comparisons were analyzed: cognitively healthy vs. MCI, cognitively healthy vs. traumatic brain injury (TBI), cognitively healthy vs. mild dementia, and cognitively healthy vs. the combined group of MCI, TBI, and mild dementia patients. Analyses were restricted to MCI, TBI, and mild dementia as memory impairment is a hallmark of these but not necessarily associated with the other abnormal diagnoses in Table 1. Similarly, analyses of the Motor Skills index score were restricted to Parkinson's disease (PD) as motor impairment is characteristic of PD but not the other abnormal diagnoses. GCS performance for all abnormal diagnoses together (*N*=421) was compared with that of cognitively healthy individuals.

For each comparison, between-groups difference was tested and false positive (Type I error) and false negative (Type II error) rate (*p*[FP], *p*[FN]) computed

Table 4. False positive rate (*p*[FP]), false negative rate (*p*[FN]), and their difference (|*p*[FP]-*p*[FN]|) at each of six cutoffs for the Mindstreams index scores across cognitive diagnoses anticipated to evidence impairment (Table 3, shaded cells) and for the Global Cognitive Score (GCS) across all cognitive diagnoses.

		Cutoff (SD units)					
Mindstreams™ Measure		**-1**	**-0.75**	**-0.5**	**-0.25**	**0**	**0.25**
Memory	*p*(FP)	0.12	0.16	0.20	0.26	0.38	0.52
	p(FN)	0.48	0.42	0.35	0.30	0.25	0.15
	p(FP)-*p*(FN)	0.36	0.26	0.15	0.04	0.13	0.37
Executive Function	*p*(FP)	0.08	0.14	0.24	0.35	0.45	0.60
	p(FN)	0.65	0.56	0.41	0.32	0.20	0.10
	p(FP)-*p*(FN)	0.57	0.42	0.16	0.02	0.25	0.50
Visual Spatial	*p*(FP)	0.11	0.18	0.26	0.32	0.44	0.53
	p(FN)	0.52	0.40	0.36	0.30	0.24	0.24
	p(FP)-*p*(FN)	0.41	0.22	0.10	0.02	0.20	0.29
Verbal Function	*p*(FP)	0.13	0.18	0.22	0.28	0.40	0.57
	p(FN)	0.49	0.37	0.23	0.12	0.07	0.02
	p(FP)-*p*(FN)	0.36	0.20	0.01	0.17	0.33	0.54
Attention	*p*(FP)	0.10	0.14	0.22	0.30	0.41	0.58
	p(FN)	0.62	0.53	0.39	0.33	0.22	0.14
	p(FP)-*p*(FN)	0.51	0.39	0.18	0.03	0.19	0.44
Information Processing Speed	*p*(FP)	0.10	0.16	0.27	0.40	0.52	0.63
	p(FN)	0.50	0.39	0.32	0.26	0.22	0.13
	p(FP)-*p*(FN)	0.40	0.23	0.05	0.13	0.30	0.50
Motor Skills	*p*(FP)	0.08	0.13	0.20	0.26	0.40	0.55
	p(FN)	0.77	0.67	0.53	0.43	0.27	0.13
	p(FP)-*p*(FN)	0.69	0.54	0.34	0.17	0.14	0.41
GLOBAL COGNITIVE SCORE	*p*(FP)	0.05	0.11	0.18	0.29	0.44	0.63
	p(FN)	0.67	0.57	0.47	0.33	0.19	0.07
	p(FP)-*p*(FN)	0.62	0.46	0.29	0.04	0.25	0.55

SD: standard deviation
Shading indicates cutoffs meeting one of the following criteria: *p*(FP) closest to 0.10, *p*(FN) closest to 0.10, or smallest |*p*(FP)-*p*(FN)|.

at six normal/abnormal cutoffs relative to the normative mean: -1, -0.75, -0.5, -0.25, 0, and +0.25 SD units. The cutoff satisfying each of the following criteria was identified:

Criterion	Condition
a	*p*(FP) closest to 0.10
b	*p*(FN) closest to 0.10
c	|*p*(FP)-*p*(FN)| smallest

Each criterion reflects a different balance between severity of FPs and FNs. For criterion *a*, the relative severity of a FN is low, for criterion *b* the relative severity of a FP is low, and for criterion *c* the relative severity of a FN and a FP is equivalent. Criterion *c* was designed to identify a best-balance normal/abnormal cutoff, criterion *a* to further parse the abnormal range into probable abnormal and abnormal sub-ranges , and criterion *b* to further parse the normal range into probable normal and normal sub-ranges. For each of the criteria, a cutoff was adopted for clinical use if criterion was met at that cutoff for the majority of comparisons tested.

As an example, for the cognitively healthy vs. MCI comparison on the Memory index score, computation of *p*(FP) an *p*(FN) for the -1SD cutoff was as follows.

A 2 x 2 table was constructed as below,

		Memory Index Score (-1SD cutoff)		
		Abnormal	Normal	TOTAL
Expert Diagnosis (standard)	Abnormal	60	68	128
	Normal	47	338	385
		107	406	513

where Abnormal = MCI and Normal = cognitively healthy.

If letters are assigned to each of the cells as follows,

		Memory Index Score (-1SD cutoff)	
		Abnormal	Normal
Expert Diagnosis (standard)	Abnormal	A	B
	Normal	C	D

p(FP)=C/(C+D) and p(FN)=B/(A+B). Substituting the actual values gives p(FP)=47/(47+338)=0.12 and p(FN)=68/(60+68)=0.53. Using these values, |p(FP)-p(FN)|=|0.12-0.53|=0.41.

p(FN), p(FP), and |p(FP)-p(FN)| were computed in the same way for the -0.75, -0.5, -0.25, and 0 SD cutoffs to give the table below.

Cutoff (SD units)	p(FP)	p(FN)	|p(FP)-p(FN)|
-1	0.12	0.53	0.41
-0.75	0.16	0.48	0.32
-0.5	0.20	0.39	0.19
-0.25	0.26	0.33	0.07
0	0.38	0.28	0.10
+0.25	0.52	0.16	0.36

This table was then examined to determine the cutoff at which each criterion was met (shaded cells). Across cutoffs, criterion *a* was met at -1SD as 0.12 is the p(FP) value closest to 0.10. Criterion *b* was met at +0.25SD, and criterion *c* was met at -0.25SD.

As indicated above, this procedure was run for each individual cell comparison and for each cross-cell comparison (Table 3). The cutoff at which each criterion was most often met was identified, separately for individual cell comparisons and for cross-cell comparisons for each index score.

Between-group tests were by independent samples t-test. If heterogeneity of variance was indicated by a significant Levene's test, the non-parametric Mann-Whitney U test was used instead. Two-tailed statistics were used throughout, and p<0.05 was considered significant. All statistics were computed with SPSS statistical software (SPSS, Chicago, IL).

Normalization

Normalization was according to a normative sample consisting of 483 participants with an expert diagnosis of cognitively healthy in controlled research studies (including those of the Neurocognitive Study Group*) using Mindstreams. Of the 401 cognitively healthy individuals in the present analysis, 383 were also part of the normative sample.

Data was normalized according to the following stratifications:

Age Group	Years of Education	N [1]
≤18	≤12	59
	>12	
>18 and ≤50	≤12	40
	>12	114
>50 and ≤70	≤12	49
	>12	89
>70	≤12	45
	>12	85

[1] maximum across all outcome parameters.

For the expanded Go-NoGo test, a relatively new test, normalization was according to a normative sample consisting of 66 cognitively healthy (mean age: 22.7±5.5 years; mean education: 11.8±2.8 years) participants in Neurocognitive Study Group* research studies. A total of 116 participants (mean age: 24.3±7.5 years; mean education: 12.3±3.3 years) in the present analysis received the expanded Go-NoGo test. Of these, 44 participants were cognitively healthy, all of whom were part of the normative sample.

Expanded Go-NoGo test data was normalized according to the following stratifications:

Age Group	N [1]
≤23	25
>23	41

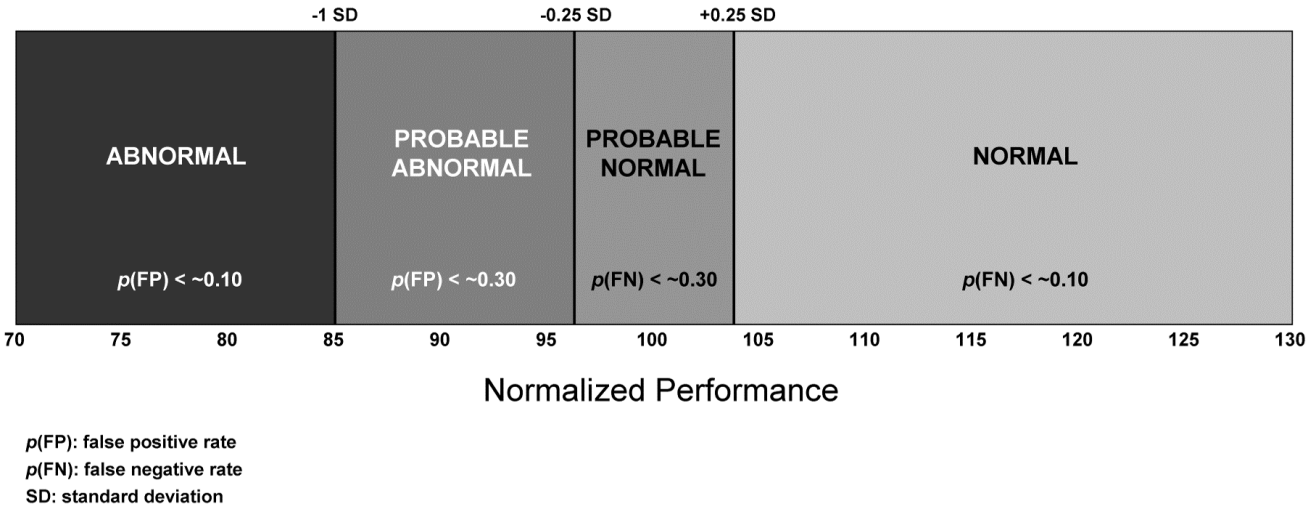
[1] maximum across all outcome parameters.

In the event of a failed Mindstreams practice session, a score equivalent to 2 percentile units was assigned. This score was also assigned for performance index outcome parameters (see above) in the event of 0% accuracy on the actual test. To limit the influence of extreme outliers, actual test performance of poorer than -4SD was replaced with the normalized score for -4SD.

Results

Robust (p<0.001) between-group differences were found for the vast majority of comparisons.

Figure 1. Revised Mindstreams Summary Measure Graph.



Individual Cell Comparisons

Criterion a: p(FP) closest to 0.10

For all 21 individual cell comparisons (Table 3), *p*(FP) was closest to 0.10 at -1 SD units.

Criterion b: p(FN) closest to 0.10

For 16 of the 21 individual cell comparisons, *p*(FN) was closest to 0.10 at +0.25SD. Criterion *b* was met at -0.75SD for 1 comparison (Memory, mild dementia), at -0.25SD for 3 comparisons (Executive Function, mild dementia; Verbal Function, mild dementia; Attention, schizophrenia), and at 0SD for one comparison (Executive Function, schizophrenia). Notably, at a cutoff of +0.25SD, *p*(FN) was less than 0.10 for these 8 comparisons.

Criterion c: |p(FP)-p(FN)| smallest

For 13 of the 21 individual cell comparisons, |*p*(FP)-*p*(FN)| was smallest at -0.25SD. Criterion *c* was met at -0.75SD for 2 comparisons (Memory, mild dementia; Attention, schizophrenia), at -0.5 SD units for 4 comparisons (Executive Function, mild dementia; Verbal Function, mild dementia; Information Processing Speed, TBI; Information Processing Speed, mild dementia), and at 0 for two comparisons (Attention, PD; Attention, HLGD).

Cross-Cell Index Score and GCS Comparisons

Criterion a: p(FP) closest to 0.10

For all 7 index score comparisons across abnormal diagnoses predicted to evidence a performance decrement (Table 3), *p*(FP) was closest to 0.10 at -1SD units (Table 4). For the GCS comparison across all abnormal diagnoses, criterion *a* was met at -0.75SD units. At a cutoff of -1SD, *p*(FP) was less than 0.10 for this comparison.

Criterion b: p(FN) closest to 0.10

For 6 of the 7 index score comparisons, *p*(FN) was closest to 0.10 at +0.25SD. Criterion *b* was met at -0.25SD for the Verbal Function index score comparison. At a cutoff of +0.25SD, *p*(FN) was less than 0.10 for this comparison. For the GCS comparison, *p*(FN) was closest to 0.10 at +0.25SD.

Criterion c: |p(FP)-p(FN)| smallest

For 4 of the 7 index score comparisons, |*p*(FP)-*p*(FN)| was smallest at -0.25SD. Criterion *c* was met at -0.5SD for the Verbal Function and Information Processing Speed index score comparisons. |*p*(FP)-*p*(FN)| was smallest at 0SD for the Motor Skills index score comparison. For the GCS comparison, criterion *c* was met at -0.25SD.

Discussion

The present analysis identifies -0.25 SD units (i.e., 96.25 normalized units) as a best-balance normal/abnormal cutoff, with equivalent severity of Type I and Type II errors (criterion *c*; Figure 1). Across a range of abnormal diagnoses and Mindstreams summary measures anticipated to

evidence impairment for those diagnoses, *p*(FP) and *p*(FN) were approximately equivalent at -0.25SD. Given this cutoff, a score above 96.25 would be considered 'normal' and a score 96.25 or below 'abnormal'. While balanced at a cutoff of -0.25SD, *p*(FP) and *p*(FN) were approximately 0.30. Hence, using this cutoff, a sizeable proportion of classifications in either the 'normal' or 'abnormal' range may be erroneous. Therefore, additional cutoffs were identified to reduce *p*(FP) in the 'abnormal' range (criterion *a*) and *p*(FN) in the 'normal' range (criterion *b*).

Across comparisons *p*(FP) was reduced to approximately 0.10 (criterion *a*) at a cutoff of -1SD (i.e., 85 normalized units). Thus this cutoff was adopted to distinguish between 'abnormal' and 'probable abnormal' (Figure 1). A score 85 or below would be considered 'abnormal', and a score from 96.25 to 85 would be considered 'probable abnormal'. Using this cutoff, there would be only very few erroneous classifications in the 'abnormal' sub-range. *p*(FN) is sizeable at -1SD (Table 4), but rather than 'normal', scores immediately above 85 are classified as 'probable abnormal' on the basis of the -0.25SD cutoff (criterion *c*). Hence scores above 96.25 are not 'abnormal', but neither are they 'normal'. Rather, they are 'probable abnormal', a designation that aptly reflects the certainty of scores in this sub-range on the basis of *p*(FP) for criterion *c*.

Across comparisons, *p*(FN) was reduced to approximately 0.10 (criterion *b*) at a cutoff of +0.25SD (i.e., 103.75 normalized units). Thus this cutoff was adopted to distinguish between 'normal' and 'probable normal'. A score above 103.75 would be considered 'normal', and a score from 96.25 through 103.75 would be considered 'probable normal'. With this cutoff, there would be hardly any misclassifications in the 'normal' sub-range. *p*(FP) is considerable at +0.25SD (Table 4), but rather than 'abnormal', scores 96.25 and immediately below are classified as 'probable normal' on the basis of the -0.25SD cutoff (criterion *c*). Hence scores 96.25 and below are not 'normal', but neither are they 'abnormal'. Rather, they are 'probable 'normal', a designation indicative of the certainty of scores in this sub-range on the basis of *p*(FN) for criterion *c*.

Taken together, the sub-ranges defined by the present analysis constitute a powerful clinical tool. Rather than simply defining a 'normal range', the present analysis defines a set of clinically relevant sub-ranges on the basis of relative error rates. The analysis utilizes an appropriate normative sample, drawn from the same controlled research studies as the cognitively impaired participants and including many of the cognitively healthy participants. Further, by defining multiple cutoffs and sub-ranges, each reflecting a different relative weighting of severity of false positives and false negatives, the analysis moves beyond the limitations of a traditional single-cutoff approach. Finally, as it based upon multiple abnormal diagnoses including many associated with only mild impairment, the set of sub-ranges is both general and sensitive for detection of impairment in varied clinical contexts.

The convention of using -1SD to -2SD as a normal/abnormal cutoff on neuropsychological tests (Spreen & Strauss, 1998) and particularly for MCI detection (e.g., Petersen et al., 1999) may be analogous to the 'probable abnormal'/'abnormal' cutoff of -1SD (criterion *a*) in the present analysis. This cutoff is associated with a relatively low severity of false negatives and is consistent with objective deficit being only one criterion for MCI and part of a more extensive workup. With the set of sub-ranges defined by the present analysis, a score immediately above -1SD is classified not as 'normal', but rather as 'probable abnormal', thus giving the clinician a more accurate picture of the cognitive status of the patient.

Alternatively, the conventional normal/abnormal cutoff of -1SD to -2SD may actually be analogous to the Mindstreams normal/abnormal cutoff of -0.25SD. If so, the disparity between the two may be attributable to the loose definition of cognitively normal in traditional normative samples and their questionable suitability as reference groups for experimental research data. Indeed the present analysis employed a strict definition of cognitively healthy both for the normative sample and for the cognitively healthy group in the study sample. This served to ensure that only individuals who were truly cognitively healthy were part of these groups and, as the same criteria were applied to both groups, to optimize the congruity between them. Given the more rigorous definition of 'normal' and the greater correspondence between normative and study samples, the normal/abnormal cutoff is higher than for a research study employing a typical neuropsychological test and the probable normal zone tightly straddles the normative mean.

The present analysis was designed for wide

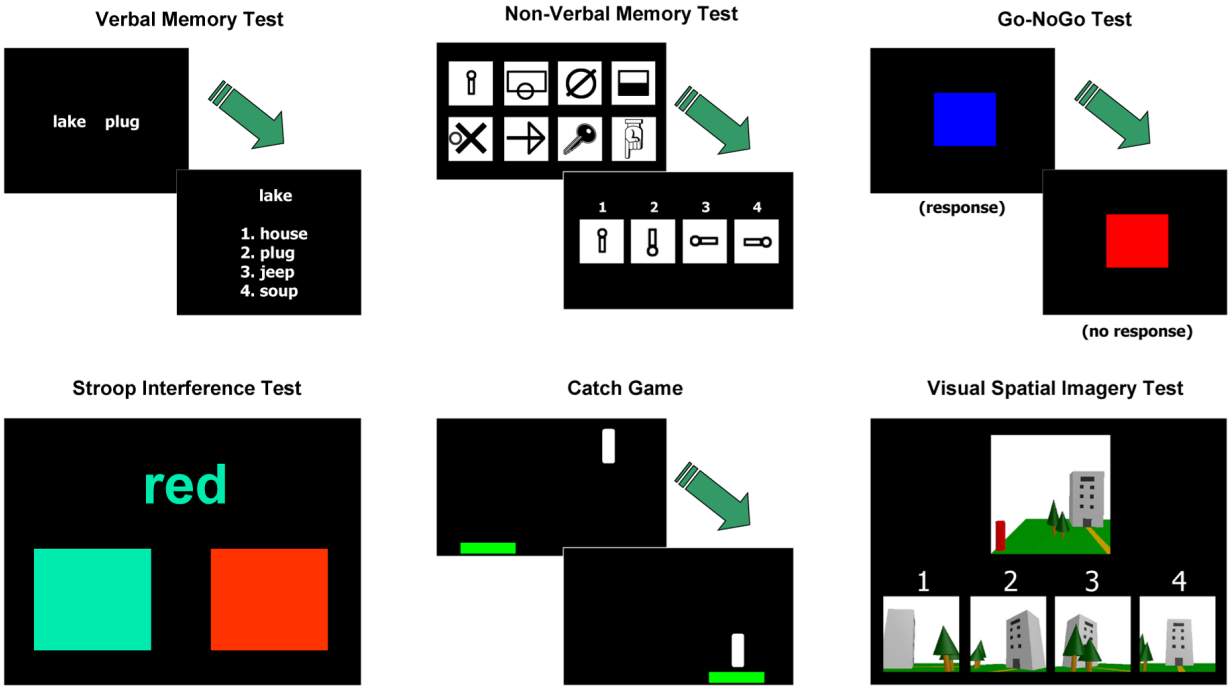
applicability across cognitive diagnoses and Mindstreams summary measures. However, it is clear that the single set of cutoffs derived herein may not be ideal for all cognitive diagnoses and summary measures. Future analyses may therefore derive diagnosis- and index-score specific cutoffs to further improve clinical utility in settings with focused applications.

Acknowledgements

Gratitude to Shimon Amit and Judy Simon for expert technical support. Funding provided by the Institute for the Study of Aging, a non-profit foundation supported by the Estee Lauder Trust.

*The Neurocognitive Study Group (in alphabetical order):

H. Chertkow (McGill-Jewish General Hospital, Montreal, Canada)
H. Crystal (State University of New York, Brooklyn, NY)
T. Dwolatzky (Ben-Gurion University of the Negev, Beer Sheva, Israel)
D. Elstein (Shaare Zedek Medical Center, Jerusalem, Israel)
N. Giladi (Tel Aviv Sourasky Medical Center, Tel Aviv, Israel)
F. Goldstein (Emory University, Atlanta, GA)
J. Hausdorff (Tel Aviv Sourasky Medical Center, Tel Aviv, Israel)
J. Lah (Emory University, Atlanta, GA)
A. Levey (Emory University, Atlanta, GA)
A. Schweiger (Academic College of Tel-Aviv, Tel Aviv, Israel; Center for Cognition and Communication, New York, NY)
R. Strous (Beer Yaakov Psychiatric Hospital, Tel Aviv University, Zrifin, Israel)
A. Zimran (Shaare Zedek Medical Center, Jerusalem, Israel)



Screenshots are adaptations of screens presented during actual testing and are provided for illustration purposes only.