Clinical Inference in the Assessment of Mental Residual Functional Capacity

David J Schretlen, PhD, ABPP
OIDAP Panel Meeting
10 June 2009
Methods of Inference

1. Pathognomonic sign approach
2. Pattern analysis
3. Level of performance or deficit measurement
**Pathognomonic Signs**

- Characteristic of particular disease or condition
- High specificity
- Present vs. absent
- Often ignored questions
  - How frequent are they in healthy individuals?
  - How reliable are they?
- 10 physicians (5 neurologists & and 5 others)
- Examined both feet of 10 participants
  - 9 w/ upper motor neuron lesions (8 unilateral; 1 bilateral)
  - 1 w/ no upper motor neuron lesion
- Babinski present in
  - 35 of 100 examinations of foot w/ UMN weakness (sensitivity)
  - 23 of 99 examinations of foot w/o UMN weakness (specificity)

*Neurology* (2005)
Fig. 4.8 The Complex Figure of Rey (Rey, 1959). Courtesy of Les Éditions du Centre de Psychologie Appliquée.
Pathognomonic?

91-year-old Caucasian woman
14 years of educ (AA degree)
Excellent health
Rx: Floxin, vitamins
MMSE = 27/30
WAIS-R MOANS IQ = 109
Benton FRT = 22/27
WMS-R VR Immed. SS = 8
Jan. 2004: 68-year-old retired engineer with reduced arm swing, bradyphrenia & stooped posture. Diagnosed with atypical PD.

Apr. 2005: Returns for follow-up testing 2 months after CABG; thinks his memory has declined slightly but PD is no worse.

Jan. 2007: Returns & wife reports visual hallucinations, thrashing in sleep, & further memory ↓ but his PD is no worse and he still drives.
Pathognomonic Signs: Limitations & Implications

- Are there any in clinical neuropsychology?
  - Unclear if there are any for a specific disease or condition
- Might be more prevalent in normal population than commonly thought
- Reliability is rarely assessed

- If we recommend that SSA rely on pathognomonic signs of impairment, we should not assume that successful job incumbents are free of such signs
Methods of Inference

1. Pathognomonic sign approach
2. Pattern analysis
3. Level of performance or deficit measurement
Pattern Analysis

- Recognizable gestalt of signs, symptoms, history, laboratory findings, and test results
- Most elaborate approach to inference
- Best for patients with typical presentations
Empirical Basis of Pattern Analysis

- Considerable empirical support
  - But much of it is pieced together from disparate studies

- Studies often involve discriminant function analyses
  - Other designs have been used (eg, comparing AD and HD patients on MMSE after matching for total score)
Examining the range of normal intraindividual variability in neuropsychological test performance

- Derived 32 z-transformed test scores for 197 healthy Ss
- Subtracted each person’s lowest z-score from his or her own highest z-score to measure the “Maximum Difference” (MD)
- Resulting MD scores ranged from 1.6 - 6.1 ($M=3.4$)
- 65% produced MD scores $\geq 3.0$; 20% had MDs $\geq 4.0$
- Eliminating each persons’ single highest and lowest test scores decreased their MDs, but 27% still produced MS values of 3.0 or greater
Intra-individual variability shown by 197 healthy adults
Pattern Analysis: Limitations & Implications

- Applicability varies with typicality of patient
- Normal variation can be mistaken for meaningful patterns
- This approach probably mirrors the task of linking specific residual functional capacities to job demands more closely than the others
- It might be useful to think about linking specific RFCs to job demands using such statistical methods as cluster analysis or canonical correlation
**Methods of Inference**

1. Pathognomonic sign approach
2. Pattern analysis
3. Level of performance or deficit measurement
Level of Performance

- Often used to detect impairments or deficits

- **But, what is an impairment or deficit?**
  - Deficient ability compared to normal peers?
  - Decline for individual (but normal for peers)?
Level of Performance: Deficit Measurement

- We infer *ability* from *performance*
  - But factors other than disease (eg, effort) can uncouple them
  - There is no one-to-one relationship between brain dysfunction and abnormal test performance *at any level*

- But even if other factors do not uncouple them, what is an *abnormal* level of performance?

- **Thought experiment**: Suppose we test the IQs of 1,000,000 perfectly healthy adults
Would the distribution look like this?

Mean: 100, Std.Dev: 15
Probably not
More likely, the distribution would be shifted up
Consequently

- If a distribution of one million IQ test scores is shifted up **10 points**, but remains Gaussian, then 4800 people will still score below **70**

- How do we understand normal, healthy people with IQs below 70?
  - Chance?
  - Healthy but nonspecifically poor specimens?
Logical Conclusions

- Some of those who perform in the lowest 2% of the distribution are *normal*.
- Most of those who perform in the lowest 2% of the distribution are *impaired*.
- The probability of impairment increases with distance below the population mean.
Cutoff Scores

- Help decide whether performance is abnormal
- Often set at 2 sd below mean, but 1.5 and even 1 sd below mean have been used
- If test scores are normally distributed, these cutoffs will include 2.3% to 15.9% of normal individuals on any single measure
Multiple Measures

- When a test battery includes multiple measures, the number of normal healthy individuals who produce abnormal scores increases.
- So does the number of abnormal scores they produce.
- Using multiple measures complicates the interpretation of abnormal performance on test batteries.
The binomial distribution can be used to predict how many abnormal scores healthy persons will produce on batteries of various lengths.

Probability of obtaining 2 or more “impaired” scores based on selected cut-off criteria & number of tests administered

<table>
<thead>
<tr>
<th>Cut-off</th>
<th>Number of Tests Administered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>--1.0 SD</td>
<td>.50</td>
</tr>
<tr>
<td>--1.5 SD</td>
<td>.14</td>
</tr>
<tr>
<td>--2.0 SD</td>
<td>.03</td>
</tr>
</tbody>
</table>

Ingraham & Aiken (1996)
Participants

- 327 reasonably healthy adults without current psychiatric illness or substance abuse/dependence

Procedure

- Administered 25 cognitive measures; obtained T-scores
- Classified T-scores as normal or “abnormal” based on three cutoffs: <40, <35, and <30
- Computed Cognitive Impairment Indices (CII) as the number of abnormal scores each person produced
- Used both unadjusted and demographically adjusted scores
We estimated how many individuals would produce 2 or more abnormal scores using three T-score cutoffs

1. Based on binomial distribution (BN)
2. Based on Monte Carlo simulation (MC) using unadjusted T-scores
3. Based on Monte Carlo simulation (MC_{adj}) using adjusted T-scores
<table>
<thead>
<tr>
<th>Test/Measure</th>
<th>M ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mini-Mental State Exam</td>
<td>28.1 ± 1.7</td>
</tr>
<tr>
<td>Grooved Pegboard Test</td>
<td></td>
</tr>
<tr>
<td>Dominant hand</td>
<td>80.4 ± 28.1</td>
</tr>
<tr>
<td>Non-dom hand</td>
<td>90.5 ± 34.7</td>
</tr>
<tr>
<td>Perceptual Comparison Test</td>
<td>64.5 ± 16.4</td>
</tr>
<tr>
<td>Trail Making Test</td>
<td></td>
</tr>
<tr>
<td>Part A</td>
<td>34.9 ± 17.0</td>
</tr>
<tr>
<td>Part B</td>
<td>95.0 ± 69.4</td>
</tr>
<tr>
<td>Brief Test of Attention</td>
<td>15.4 ± 3.7</td>
</tr>
<tr>
<td>Modified WCST</td>
<td></td>
</tr>
<tr>
<td>Category sorts</td>
<td>5.3 ± 1.3</td>
</tr>
<tr>
<td>Perseverative errors</td>
<td>2.5 ± 3.9</td>
</tr>
<tr>
<td>Verbal Fluency</td>
<td></td>
</tr>
<tr>
<td>Letters cued</td>
<td>28.2 ± 9.2</td>
</tr>
<tr>
<td>Category cued</td>
<td>44.8 ± 11.4</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>28.2 ± 2.6</td>
</tr>
<tr>
<td>Benton Facial Recognition</td>
<td>22.4 ± 2.3</td>
</tr>
<tr>
<td>Rey Complex Figure</td>
<td>31.3 ± 4.3</td>
</tr>
<tr>
<td>Clock Drawing</td>
<td>9.5 ± 0.8</td>
</tr>
<tr>
<td>Design Fluency Test</td>
<td>14.2 ± 7.2</td>
</tr>
<tr>
<td>Wechsler Memory Scale</td>
<td></td>
</tr>
<tr>
<td>Logical Memory I</td>
<td>26.3 ± 6.9</td>
</tr>
<tr>
<td>Logical Memory II</td>
<td>22.4 ± 7.5</td>
</tr>
<tr>
<td>Hopkins Verbal Learning Test</td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td>24.6 ± 4.8</td>
</tr>
<tr>
<td>Delayed recall</td>
<td>8.7 ± 2.6</td>
</tr>
<tr>
<td>Delayed recognition</td>
<td>10.4 ± 1.6</td>
</tr>
<tr>
<td>Brief Visuospatial Memory Test</td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td>22.2 ± 7.5</td>
</tr>
<tr>
<td>Delayed recall</td>
<td>8.7 ± 2.7</td>
</tr>
<tr>
<td>Delayed recognition</td>
<td>5.6 ± 0.7</td>
</tr>
<tr>
<td>Prospective Memory Test</td>
<td>0.6 ± 0.7</td>
</tr>
</tbody>
</table>
25 Measure Battery

Predicted and observed percentages of participants who produced 2 or more abnormal test scores (y axis) as defined by three different cutoffs (<40, <35, and <30 T-score points)
Spearman correlations between Cog Imp Index scores based on unadjusted T-scores and age, sex, race, years of education and estimated premorbid IQ

<table>
<thead>
<tr>
<th>No. of tests</th>
<th>T-score cutoff</th>
<th>Mean (SD)</th>
<th>Age</th>
<th>Sex</th>
<th>Race</th>
<th>Educ.</th>
<th>NART IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>&lt; 40</td>
<td>3.6 (4.4)</td>
<td>.573**</td>
<td>-.029</td>
<td>.215**</td>
<td>-.327**</td>
<td>-.360**</td>
</tr>
<tr>
<td>25</td>
<td>&lt; 35</td>
<td>1.6 (2.7)</td>
<td>.528**</td>
<td>-.039</td>
<td>.186*</td>
<td>-.325**</td>
<td>-.354**</td>
</tr>
<tr>
<td>25</td>
<td>&lt; 30</td>
<td>0.5 (1.3)</td>
<td>.409**</td>
<td>-.066</td>
<td>.176</td>
<td>-.312**</td>
<td>-.318**</td>
</tr>
</tbody>
</table>

* = p < 0.001; ** = p < 0.0001
This study shows that

- Neurologically normal adults produce abnormal test scores
  - Rate varies with battery length & cutoff used to define abnormal

- This is not due purely to chance
  - Varies with age, education, sex, race and est. premorbid IQ
  - Demographically adjusting scores eliminates the relationship between these characteristics and abnormal performance

- Findings underscore distinction between “abnormal” test performance and “impaired” functioning
  - Test performance can be abnormal for many reasons: impaired functioning is but one
Returning to the question of what cut-off we should use to define abnormal performance...

- Stringent cut-offs decrease test sensitivity
- Liberal cut-offs decrease test specificity
- Adding tests increases the risk of type I errors
- Excluding tests increases the risk of type II error

- As in most endeavors, we must exercise judgment
Decline from Premorbid Ability

- If we know a person’s “premorbid” ability, then it is relatively simple to determine decline
  - Unfortunately, we rarely know this
  - Therefore, we have to estimate it
  - So how do we do that?

- Research has focused on estimating premorbid IQ
Estimating Premorbid IQ

- Demographic prediction
  - Barona formula $SE_{est} = 12$ points (95% CI = ±24 points)

- Word reading tests are more accurate
  - Except for persons with very limited education
  - And those with aphasia, reading disorders, or severe dementia
  - And persons for whom English is a second language
Stability of NART-R IQ Estimates

![Graph showing the stability of NART-R IQ estimates over a 5-year follow-up period. The Rsq value is 0.9479.](image)
Correlation of NART-R and WAIS-R

Current Est. FSIQ

NART IQ

$R^2 = 0.5776$
But how well does the NART-R predict cognitive abilities other than IQ?

Administered 26 cognitive measures to 322 healthy adults

Regressed each on age, saved the residuals, and correlated these with NART-R scores

Compared the correlation of NART-R and IQ with correlations of the NART-R and other age-adjusted cognitive measures
NART-R correlation with FSIQ = .72

NART-R correlations with other test scores ranged from - .53 to .48

(Every one of the latter was significantly smaller than the correlation with FSIQ)
Estimating Premorbid Abilities

- An *essential* and *unavoidable* aspect of every neuropsychological examination
- If we don’t do explicitly, then we do it implicitly
- Even the best methods yield ballpark estimates
- We’re better at estimating premorbid IQ than other premorbid abilities
Examined 28 scores derived from 16 cognitive tests that were administered to 221 reasonably healthy adults

Grouped participants by WAIS-R Full Scale IQ into three groups:

- **N = 37** Below average (BA)  
  FSIQ < 90  
  Mean = 83
- **N = 106** Average (A)  
  FSIQ 90-109  
  Mean = 101
- **N = 78** Above average (AA)  
  FSIQ > 109  
  Mean = 121
Intelligence and Cognitive Functioning

- Correlations between intelligence and other cognitive abilities are stronger below than above IQ scores of 110
  - *It is less likely that smart people will do well on other tests than it is that dull people will do poorly*

- A normal person with an IQ of 85 is likely to produce “impaired” scores on about 10% of other cognitive tests
Deficit Measurement: Limitations & Implications

- No isomorphic relationship between performance and ability
- Adding tests can increase false positive (type 1) errors
- Setting stringent cut-offs can increase misses (type 2) errors
- NART predicts pre-morbid IQ better than other abilities
- Raising “cut-off” scores for patients of above average IQ can compound the problem of multiple comparisons
Deficit Measurement: Limitations & Implications

Many – if not most – **successful** job incumbents likely fall short of meeting one or more of their job demands.

What cutoff in the distribution of an ability shown by successful job incumbents should we use to define sufficient RFC for someone to do that job? This will **directly** affect the percentage of applicants who will be found disabled.

Factors other than impairment, like effort, can uncouple the linkage between performance and ability.

Work demands, RFC, and “deficit” vs. “impairment”