Evidence-Based Practice and the Use of Reliable Change Methods

Gordon J. Chelune, Ph.D.
Professor of Neurology, University of Utah
Center for Alzheimer’s Care, Imaging and Research

http://www.utahmemory.org
Education tab – CE & CME courses
Look for INS Jerusalem 2014: EBM & Reliable Change
Learning Objectives:

As a result of attending this presentation, participants will be able to:

1. Discuss and explain the critical factors that affect the reliability and fidelity of serial assessments and the unique statistical features of change scores;

2. Compare and contrast reliable change methods that estimate and/or measure the dispersion of change scores and how these methods can be linked to base-rate information and Test Operating Characteristic to inform clinical practice and enhance clinical research;

3. Use simple summary data in test manuals and research reports to create regression equations to evaluate the significance between observed and predicted retest scores.
Themes

Evidence-Based Practice
Clinically Meaningful Change
A Tool for Assessing Change
What is Evidence-Based Clinical Neuropsychology?

A value-driven pattern of clinical practice that attempts to integrate “best research” derived from the study of populations to inform clinical decisions about individuals within the context of the provider’s expertise and individual patient values with the goal of maximizing clinical outcomes and quality of life for the patient in a cost-effective manner while addressing the concerns and needs of the provider’s referral sources.

Adapted from Chelune, 2010
Clinical Practice and Evidence-Based Medicine

The impetus of evidence-based practice medicine has its roots in the “outcomes movement” of the 1980s when it became increasingly apparent to payers and many practitioners that a significant portion of health care expenditures was wasted on unproven or ineffective tests and treatments.

Horwitz, 1996
Outcomes accountability and the management of individual patients on the basis of epidemiologic information regarding outcomes became increasingly critical to the practice of medicine

Johnson, 1997
In a broad sense, clinical outcomes are discrete measurable events, marked by a change in status, performance, or other objectively defined endpoint, that can be tracked both in the aggregate on a group level but also, importantly, at the level of the specific patient.

To be useful in the care of patients, outcomes data must analyzed and packaged in such a manner that they can be directly "used" by the end-user, namely the clinician.

Chelune, 2002, 2010
Assumptions:

Every Clinical Patient Assessment...

- Represents a Clinical Outcome
- Can be interpreted within context of Evidence-based Research
One of the defining features of evidence-based practice is the use of data derived from research based on populations to inform clinical decisions about individuals.

...how do we move from group data to data that is applicable at the level of the individual?
Has there been a “change” in the patient’s clinical status?
How Do We Go About Recognizing Significant Change in Neuropsychological Performance?

And Why Is It Important?
Clinical Assessments Involve Inferences About CHANGE

- Single-Point Assessment – Does the observed test score represent a meaningful difference from an inferred premorbid?
- Serial Assessment – Does the observed retest score represent a meaningful change/difference from baseline?

It’s All About Change and the Base Rates of Change
The Role of Serial Assessments in Clinical Neuropsychology

- Evaluating recovery from an injury
- Documenting progression of a neurodegenerative process
- Examining the impact of rehabilitation or other intervention (surgical intervention, drug therapy, etc)
- Providing a routine follow-up to a previous evaluation
- Conducting forensic neuropsychological assessments
Has there been a “change” in my patient’s clinical status?
Serial Assessment

Test 1    Test 2    Test 3


Difference 1    Difference 2
“No man ever steps in the same river twice, for it's not the same river and he's not the same man.”

_Heraclitus_

535–475 BC
Reliable vs. Meaningful Change

Matarazzo & Herman: JCN, 1984

**Statistical Reliability** describes whether an observed level of difference between 2 scores is apt to be a reliable and repeatable difference from 0 and not due to measurement error and chance fluctuations in the scores.

Addresses the Question of “How Much”
Is the difference between groups statistically reliable?

$p < .05$

Performance
Reliable vs. Meaningful Change
Matarazzo & Herman: JCN, 1984

**Clinical Abnormality** describes whether an observed difference between 2 scores is sufficiently rare in a normal population that it is more likely to be obtained in a population that is external to the normative group (i.e. an “abnormal” population).

Addresses the Question of “How Many”
The Question: “How Many”

Optimal Cut-off
Maximizes Sensitivity and Specificity
Best Overall Hit Rate

Performance

COI

RP

True Positives (Sensitivity)

True Negatives (Specificity)

FP FN
Assessment of Test-Retest Change

Statistical Issues

**Stability**
- (Bias)
- Variable of Interest
  - Practice Effects
  - Demographic Influences

**Reliability**
- (Error)
- Measurement Error
  - Regression to the Mean
  - Random Events
Distribution of Difference (Change) Scores

Perfect Stability and near perfect Reliability
Distribution of Difference (Change) Scores

Perfect Stability but less than perfect Reliability
Distribution of Difference (Change) Scores

Less than perfect reliability and a systematic bias or practice effect
Regression to the Mean

\[ X_{p2} = X_{01} - [(X_{01} - M)(1 - r_{12})] \]

**Test A:**

- \( M = 100 \)  \( r_{12} = .90 \)
- \( X_{1a} = 120 \)  \( X_{p2} = 120 - [(120 - 100)(1 - .90)] = 118 \)
- \( X_{1b} = 80 \)  \( X_{p2} = 80 - [(80 - 100)(1 - .90)] = 82 \)
- \( X_{1c} = 70 \)  \( X_{p2} = 70 - [(70 - 100)(1 - .90)] = 73 \)

**Test B:**

- \( M = 100 \)  \( r_{12} = .75 \)
- \( X_{1a} = 120 \)  \( X_{p2} = 120 - [(120 - 100)(1 - .75)] = 115 \)
- \( X_{1b} = 80 \)  \( X_{p2} = 80 - [(80 - 100)(1 - .75)] = 85 \)
- \( X_{1c} = 70 \)  \( X_{p2} = 70 - [(70 - 100)(1 - .75)] = 77.5 \)
“Change” is a Variable:

Has its own Unique Statistical Properties

They are dependent on the Stability and Reliability of the assessment tool

Alternately, the degree of independence from Bias and Error
How Should We Deal With Issues of Bias and Error?

1. Alternate Forms

Four alternate forms administered 4x, 1-2 weeks apart

“Reasonable alternate form comparability was demonstrated ... Nonetheless, alternate forms are likely to be an insufficient means of controlling practice in speeded measures at brief (1-2 weeks) retest intervals. Reliable change indices demonstrated that practice must be accounted for in individual retesting.”
How Should We Deal With Issues of Bias and Error?

1. Alternate Forms
2. Extend the Test-Retest Interval
How Should We Deal With Issues of Bias and Error?

1. Alternate Forms
2. Extend the Test-Retest Interval
3. Reliable Change Methods
Reliable Change Indexes

2. Reliable Change Index (Jacobsen & Truax, 1991)
3. Practice Adjusted Reliable Change Index (Chelune, et al., 1993)
4. Modified Practice Adjusted Reliable Change Index (Iverson, 2001)
5. Reliable Change based on Standard Error of Prediction (Basso, et al., 1999; Theisen et al., 1998)
Two ways to determine the frequency or base rate of an observed score relative to a reference population:

1. Simple Difference Approach
2. Predicted Difference Approach
## Simple Difference Approach

\[
\frac{(X2 - X1)}{SD} = z\text{-score}
\]

**Standard Deviation of the Difference Scores**

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient A</td>
<td>100</td>
</tr>
<tr>
<td>Patient B</td>
<td>100</td>
</tr>
</tbody>
</table>

### WAIS-III Test-Retest Sample (N=373)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>99.85</td>
<td>13.63</td>
</tr>
<tr>
<td>Time 2</td>
<td>104.36</td>
<td>14.97</td>
</tr>
<tr>
<td>Difference</td>
<td>4.51</td>
<td><strong>4.77</strong></td>
</tr>
<tr>
<td>Test-Retest (r_{xy})</td>
<td>.949</td>
<td></td>
</tr>
</tbody>
</table>

**Reliable Change Index** methods for estimating SD\(_{diff}\)
Patient A  
\[ 95 - 100 = -5.00 \]
\[ -4.51 \]
\[ -9.51 \]

\[ z - \text{score} = \left( \frac{-9.51}{4.77} \right) = -1.99 \]
\[ *p = 2.3\% \quad \text{statistically rare} \]

*Area under the Unit Curve

Adjusted for Practice Effect
Case Example

Patient B  \[ 105 - 100 = +5.00 \]
\[ -4.51 \]
\[ +0.49 \]

\[ z - \text{score} \ (\frac{+0.49}{4.77}) = 0.10 \]

*p = 46%  ns

*Area under the Unit Curve
Estimated Reliable Change Indexes
(Jacobson & Truax, 1991)

RCI Describes a Confidence Interval Around the Mean Difference Score

Represents the standard error of the difference between two test scores and describes the spread of the distribution of changes scores that would be expected if no actual change had occurred.
Estimating the Standard Deviation of Differences: $S_{diff}$
Standard Error of the Difference

Represents the standard error of the difference between two test scores and describes the spread of the distribution of changes scores that would be expected if no actual change had occurred.

Standard Error of Measurement:

\[ SE_m = SD_1 \sqrt{1 - r_{12}} \]

Standard Error of the Difference:

\[ S_{\text{diff}} = \sqrt{2(SE_m)^2} \]

Jacobsen & Truax, 1991
$S_{\text{diff}}$ of T2–T1 Differences

Individual Change:
“...in the same spirit of Cohen’s effect size”

$$Z\text{-Score} = \frac{(T2_i - T1_i)}{S_{\text{diff}}}$$

Less the perfect reliability and a systematic bias or practice effect
### Reliable Change Index

Practice Adjusted Reliable Change Index  
(Chelune, et al., 1993)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>99.85</td>
<td>13.63</td>
</tr>
<tr>
<td>Time 2</td>
<td>104.36</td>
<td>14.97</td>
</tr>
<tr>
<td>Difference</td>
<td>4.51</td>
<td>4.77</td>
</tr>
<tr>
<td>Test-Retest $r_{xy}$</td>
<td>.949</td>
<td></td>
</tr>
</tbody>
</table>

**Standard Error of Measurement:**

$$SE_m = SD_1 \sqrt{1 - r_{12}}$$

$$= 3.078$$

**Standard Error of the Difference:**

$$S_{diff} = \sqrt{2(SE_m)^2}$$

$$= 4.353$$
Case Example

Patient A

\[ 95 - 100 = -5.00 \]

\[ \frac{-4.51}{4.35} \]

\[ -9.51 \]

\( z - \text{score} = \frac{-9.51}{4.35} = -2.19 \)

\( *p = 1.43\% \) statistically rare

*Area under the Unit Curve

Adjusted for Practice Effect

\( S_{\text{diff}} \)
Reliable Change Index

Modified Practice Adjusted Reliable Change Index
(Iverson, 2001)

WAIS-III Test-Retest Sample (N=373)

Standard Error of Measurement:

\[ SE_{m1} = SD_1 \sqrt{1 - r_{12}} = 3.078 \]
\[ SE_{m2} = SD_2 \sqrt{1 - r_{12}} = 3.381 \]

Standard Error of the Difference:

\[ S_{\text{diff}} = \sqrt{(SE_{m1} + SE_{m2})^2} = 4.571 \]
Reliable Change Index

Reliable Change based on Standard Error of Prediction
(Basso, et al., 1999; Theisen et al., 1998)

WAIS-III Test-Retest Sample (N=373)

Standard Error of the Prediction ($SE_p$)

$$SE_p = SD_2\sqrt{(1 - r_{12}^2)}$$

$$= 4.719$$
## Comparison of RCI Methods

<table>
<thead>
<tr>
<th></th>
<th>$S_{\text{diff}}$</th>
<th>z-score</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs SD of diff</td>
<td>4.77</td>
<td>-1.99</td>
<td>2.30</td>
</tr>
<tr>
<td>RCI</td>
<td>4.35</td>
<td>-2.19</td>
<td>1.43</td>
</tr>
<tr>
<td>Mod RCI</td>
<td>4.57</td>
<td>-2.17</td>
<td>1.50</td>
</tr>
<tr>
<td>$\text{SE}_p$</td>
<td>4.72</td>
<td>-2.02</td>
<td>2.17</td>
</tr>
</tbody>
</table>
Limitations of the Simple Difference Method

The mean difference score (practice effect / discrepancy) is treated as a constant.

No provision is made for the influence of regression to the mean.

Does not account for other factors that might affect the difference scores.

When the observed SD of differences is not known, RCI methods vary in how they estimate the $SD_{diff}$. 
Reliable Change Index Scores
(Simple Difference Approach)

Rest in Peace
– along with pneumoencephalopgrams, Piotrowski signs, leucotomies...
Predicted Difference Method

• Regression Equation to predict WHERE a person should be on variable Y given knowledge of variable X.
  
  Prediction Equation -- $Y_p = bX_b + C$

  Yields predicted retest scores based on baseline scores and relevant demographic characteristics

• (Observed – Predicted): How FAR away a person is from where s/he should be on Retest/Variable Y.

• Normalized Z-Scores of Change –
  
  $z$-score $= \frac{(Y_o-Y_p)}{SEE}$ (Standard Error of Estimate)

  $(Observed-Predicted)/SEE$
Regression Based Change Scores
Predicting Retest FSIQs

Baseline FSIQ
Age
Education
Sex
Test-Retest Int. → Retest FSIQ
WAIS-III FSIQ

\[ Y_p = \beta Y_b + \beta X_{\text{age}} + C \]

\[ Y_p = (1.036 \times \text{FSIQ}_1) + (-0.04155 \times \text{Age}) + 3.064 \]

\[ R = 0.951 \quad R^2 = 0.904 \]

\[ \text{SEE} = 4.6478 \]

4.77 SD_{diff}

Data with Permission © The Psychological Corporation
\( Y_p = (1.036 \times FSIQ_1) + (-0.04155 \times \text{Age}) + 3.064 \)

\( Y_p = (1.036 \times 100) + (-0.04155 \times 67) + 3.064 \)

Observed Retest = 95.00

Predicted Retest = \(-103.88\) - 8.88

z-score = \(-8.88/4.6478 = -1.91\)

*\( p = 2.81\% \) statistically rare

*Area under the Unit Curve
Advantages of the Predicted Difference Method

- Differences in Initial Baseline Scores
- Practice Effects
- Regression to the Mean
- Differences in Reliability Between Measures
- Takes into account relevant demographic variables

### WAIS–IV Demographically Adjusted Composite Score Comparisons

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Score 1</th>
<th>Score 2</th>
<th>Difference</th>
<th>Critical Value</th>
<th>Significant Difference</th>
<th>Directional Base Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCI - PRI</td>
<td>42</td>
<td>60</td>
<td>-18</td>
<td>5.89</td>
<td>Y</td>
<td>4.2%</td>
</tr>
<tr>
<td>VCI - WMI</td>
<td>42</td>
<td>62</td>
<td>-20</td>
<td>5.89</td>
<td>Y</td>
<td>2.3%</td>
</tr>
<tr>
<td>VCI - PSI</td>
<td>42</td>
<td>59</td>
<td>-17</td>
<td>8.09</td>
<td>Y</td>
<td>7.8%</td>
</tr>
<tr>
<td>PRI - WMI</td>
<td>60</td>
<td>62</td>
<td>-2</td>
<td>6.21</td>
<td>N</td>
<td>42.9%</td>
</tr>
<tr>
<td>PRI - PSI</td>
<td>60</td>
<td>59</td>
<td>1</td>
<td>8.33</td>
<td>N</td>
<td>48.0%</td>
</tr>
<tr>
<td>WMI - PSI</td>
<td>62</td>
<td>59</td>
<td>3</td>
<td>8.33</td>
<td>N</td>
<td>39.4%</td>
</tr>
<tr>
<td>FSIQ - GAI</td>
<td>56</td>
<td>51</td>
<td>5</td>
<td>2.44</td>
<td>Y</td>
<td>10.8%</td>
</tr>
<tr>
<td>GAI - VCI</td>
<td>51</td>
<td>42</td>
<td>9</td>
<td>3.04</td>
<td>Y</td>
<td>5.1%</td>
</tr>
<tr>
<td>GAI - PRI</td>
<td>51</td>
<td>60</td>
<td>-9</td>
<td>2.92</td>
<td>Y</td>
<td>5.5%</td>
</tr>
</tbody>
</table>

### WMS–IV Demographically Adjusted Index Score Comparisons

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Score 1</th>
<th>Score 2</th>
<th>Difference</th>
<th>Critical Value</th>
<th>Significant Difference</th>
<th>Directional Base Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMI - VMI</td>
<td>44</td>
<td>44</td>
<td>0</td>
<td>5.54</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>IMI - DMI</td>
<td>38</td>
<td>49</td>
<td>-11</td>
<td>6.51</td>
<td>Y</td>
<td>2.6%</td>
</tr>
</tbody>
</table>

### WAIS–IV Composite Comparisons

<table>
<thead>
<tr>
<th>Composite</th>
<th>Time 1 Actual</th>
<th>Time 2 Actual</th>
<th>Time 2 Predicted</th>
<th>Time 2 Actual-Predicted Difference</th>
<th>Critical Value</th>
<th>Significant Difference</th>
<th>Base Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCI</td>
<td>100</td>
<td>83</td>
<td>103</td>
<td>-20</td>
<td>8.11</td>
<td>Y</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>PRI</td>
<td>81</td>
<td>82</td>
<td>90</td>
<td>-8</td>
<td>8.67</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>PSI</td>
<td>59</td>
<td>53</td>
<td>67</td>
<td>-14</td>
<td>12.26</td>
<td>Y</td>
<td>2-5%</td>
</tr>
</tbody>
</table>
But .....  

I don’t have access to the standardization data to derive the regression equations and to do the computations.

Besides, doing all of those computations for each patient seems really tedious and prone to error!


http://homepages.abdn.ac.uk/j.crawford/pages/dept/

Go to Downloads

RegBuild_MR.exe: Builds a multiple regression equation and uses it to make inferences concerning a case based on simple summary data that are readily available.

Alternately, go to:

http://www.utahmemory.org

Education tab: Professional CE/CME Courses

Look for INS 2014 Jerusalem: EBM & Reliable Change
RegBuild_MR.exe:

What is needed:
1. Sample Size (N)
2. Means and Standard Deviations
3. Zero-order Correlations

Available at:
http://homepages.abdn.ac.uk/j.crawford/pages/dept/
Under Downloads
Or
http://www.utahmemory.org
Education tab: Professional CE/CME Courses
INS 2014 Jerusalem: EBM & Reliable Change
This computer program (RegBuild_MR.exe) accompanies the paper by Crawford, J.R., Garthwaite, P.H., Denham, A.K., & Chelune, G.J. Using regression equations built from summary data in the psychological assessment of the individual case: extension to multiple regression. Psychological Assessment, in press. The program builds a multiple regression equation from sample summary statistics (means, SDs, and correlation matrix). The equation is then applied to the data from the individual case. The program: (a) calculates the case’s predicted score;

- 95% credible limit required...
  - Two-sided
  - One-sided lower
  - One-sided upper

- Number of predictor (i.e., X) variables:

- Sample size [N] of sample providing the summary data:

<table>
<thead>
<tr>
<th>Var No.</th>
<th>Variable Name</th>
<th>Mean</th>
<th>SD</th>
<th>Case’s Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Criterion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Predictor 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Predictor 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Predictor 3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continue | Clear Data | Exit
This program (RegBuild_MReLexe) accompanies the paper by Crawford, J.R., Garthwaite, P.H., Schneider, W.J., Williamson, S., & Chelune, G.J. Using regression equations built from summary statistics in the psychological assessment of the individual case: Analysis of the reliability of discrepancies between obtained and predicted scores. The program is an upgraded version of an earlier program (RegBuild_MReLexe; Crawford et al., 2012). It builds a multiple regression equation from sample summary statistics [means, SDs, and correlation matrix].

User's Notes:

1) 95% limits on ABNORMALITY of discrepancy...
   - Two-sided
   - One-sided lower
   - One-sided upper

2) Analysis of RELIABILITY of discrepancy...
   - NOT required
   - Required: Reliabilities from PRIMARY sample
   - Required: Reliabilities from SECONDARY sample

3) Number of predictor [i.e., \( X \)] variables :

4) Size \( (N) \) of sample providing the summary data:

<table>
<thead>
<tr>
<th>Var No.</th>
<th>Variable Name</th>
<th>Mean</th>
<th>SD</th>
<th>Case's Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Criterion</td>
<td>104.36</td>
<td>14.97</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>Predictor 1</td>
<td>99.85</td>
<td>13.63</td>
<td>100</td>
</tr>
</tbody>
</table>

Continue  Clear Data  Exit
THIS PROGRAM IMPLEMENTS METHODS DEVELOPED IN THE FOLLOWING PAPERS:


INPUTS:
Number of predictor (i.e., X) variables = 1
Sample size (n) for sample providing the summary data = 373
Analysis option: Required confidence interval on abnormality of discrepancy: Two-sided
Analysis option: Analysis of the RELIABILITY of the discrepancy is: Not required

INPUTS: Summary statistics for sample providing the data together with scores of the case:

<table>
<thead>
<tr>
<th>Measure</th>
<th>Control Mean</th>
<th>Control SD</th>
<th>Scores for Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion</td>
<td>104.360</td>
<td>14.970</td>
<td>95.000</td>
</tr>
<tr>
<td>Predictor 1</td>
<td>99.850</td>
<td>13.630</td>
<td>100.000</td>
</tr>
</tbody>
</table>

INPUTS: Correlation(s) between criterion and predictor(s) in sample:
1.000000   0.943000
\[ Y' = (1.042 \times \text{FSIQ1}) + 0.286 \]

SEE

\[ z\text{-score} = \frac{-9.5163}{4.7324} \]
(E) ANALYSIS OF THE INDIVIDUAL CASE: ABNORMALITY of discrepancy

Estimation error confidence interval (95%) for PREDICTED score = 104.5163 (95.211 to 113.822)

Effect size (Z-DF) for DISCREPANCY (plus 95% CI) = -2.016 (-2.190 to -1.834)

Significance test (t) on the discrepancy between the case's obtained and predicted scores:

<table>
<thead>
<tr>
<th>t value (on 371 df)</th>
<th>One-tailed probability</th>
<th>Two-tailed probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.0109</td>
<td>0.0225</td>
<td>0.0451</td>
</tr>
</tbody>
</table>

Estimated percentage of population obtaining a discrepancy more extreme than the case = \( \frac{2.252928\%}{\text{NS}} \)

95% confidence limits on the percentage = 1.42544% to 3.3160%
Predicting Retest FSIQs

Baseline FSIQ
Age
Education
Sex
Test-Retest Int. → Retest FSIQ
There is a wealth of data in the published research literature to which these methods can be applied.
Assessing Reliable Cognitive Decline in Older Adults: Part I
Efficacy of Multivariate Regression Equations Derived from Summary Data

Gordon Chelune, 1 John Crawford, 2 John Sheehan, 1 Kevin Duff, 1 James Holdnack, 3 Otto Pedraza *4
1University of Utah School of Medicine, 2University of Aberdeen, 3Pearson, 4Mayo Clinic

<table>
<thead>
<tr>
<th>MOANS Variable</th>
<th>Means, Standard Deviations, and Sample Sizes</th>
<th>Zero Order Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M/SD</td>
</tr>
<tr>
<td>BNT</td>
<td>1155</td>
<td>M</td>
</tr>
<tr>
<td>DRS</td>
<td>1050</td>
<td>M</td>
</tr>
<tr>
<td>COWA</td>
<td>582</td>
<td>M</td>
</tr>
<tr>
<td>AVF</td>
<td>359</td>
<td>M</td>
</tr>
<tr>
<td>Trails A</td>
<td>329</td>
<td>M</td>
</tr>
<tr>
<td>Trails B</td>
<td>329</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This computer program (RegBuild_MR.exe) accompanies the paper by Crawford, J.R., Garthwaite, P.H., Denham, A.K., & Chelune, G.J. Using regression equations built from summary data in the psychological assessment of the individual case: extension to multiple regression. The program builds a multiple regression equation from sample summary statistics (means, SDs, and correlation matrix). The equation is then applied to the data from the individual case. The program builds a multiple regression equation from sample summary statistics (means, SDs, and correlation matrix).

User’s Notes: Equation for MOANS BNT Test-Retest

95% credible limit required...
- Two-sided
  - One-sided lower
  - One-sided upper

Number of predictor (i.e., X) variables: 3

Sample size (N) of control sample: 1155

<table>
<thead>
<tr>
<th>Var No.</th>
<th>Name</th>
<th>Mean</th>
<th>SD</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BNT2</td>
<td>53.2</td>
<td>5.7</td>
<td>51</td>
</tr>
<tr>
<td>2</td>
<td>BNT1</td>
<td>53.67</td>
<td>5.8</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>Age</td>
<td>76.0</td>
<td>8.0</td>
<td>69</td>
</tr>
<tr>
<td>4</td>
<td>Educ</td>
<td>14.3</td>
<td>2.9</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.865</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.339</td>
<td>-0.397</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.35</td>
<td>3.55</td>
<td>-0.051</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Beta Weights

\[ Y' = 0.828 \times \text{BNT1} - 0.001 \times \text{Age} + 0.07 \times \text{Edu} + 7.832 \]
RESULTS FROM ANALYSIS OF THE INDIVIDUAL CASE:

Case’s OBTAINED score on Task of Interest = 51.0000
Case’s PREDICTED score from regression equation = 55.1865
Discrepancy (obtained minus predicted) between case’s obtained and predicted scores = -4.1865

Effect size (Z-OP) for discrepancy between obtained and predicted scores (plus 95% CI):
Effect size (Z-OP) = -1.424 (95% CI = -1.510 to -1.325)

Standard error for an additional (i.e., N+1th) case = 2.9468

Significance test (t) on the discrepancy between the case’s obtained and predicted scores:
t value (on 1151 df) = -1.4207
One-tailed probability = 0.0778
Two-tailed probability = 0.1557

Estimated percentage of population obtaining a discrepancy more extreme than the case = 7.783336%
95% confidence limits on the percentage = 6.4470% to 9.2552%
How Many? -- Bayesian approach: Analyses of Changes in Base Rates

Bayes’ Theorem: What we know after giving a test in equal to what we knew before doing the test times a modifier (based on the test results). Test results are used to adjust a prior distribution to form a new posterior distribution of scores.

Value Driven Pattern of Practice
The Basic 2x2 Table

<table>
<thead>
<tr>
<th>Condition of Interest</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>False Positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>B</td>
<td>D</td>
</tr>
<tr>
<td>True Negative</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factor (event)

Yes +

No -
## Bayesian Test Operating Characteristics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>% Prevalence</td>
<td>Odds</td>
</tr>
<tr>
<td>% Overall Correct Hit Rate</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>Relative Risk Ratio</td>
</tr>
<tr>
<td>Specificity</td>
<td>Likelihood Ratio</td>
</tr>
<tr>
<td>Positive Predictive Power</td>
<td>Pre – Post Test Odds</td>
</tr>
<tr>
<td>Negative Predictive Power</td>
<td>Pre – Post Test Probabilities</td>
</tr>
</tbody>
</table>

Chelune, TCN 2010
<table>
<thead>
<tr>
<th></th>
<th>Total Sample (N=383)</th>
<th>Mild Impairment MMSE 18-23 (n=98)</th>
<th>Normal Cognition MMSE &gt; 24 (n=285)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Age&lt;sup&gt;a&lt;/sup&gt;</td>
<td>72.3</td>
<td>8.1</td>
<td>74.3</td>
</tr>
<tr>
<td>Education</td>
<td>15.0</td>
<td>2.7</td>
<td>14.9</td>
</tr>
<tr>
<td>MMSE&lt;sup&gt;b&lt;/sup&gt;</td>
<td>25.6</td>
<td>3.1</td>
<td>21.1</td>
</tr>
<tr>
<td>Retest Interval (mos.)</td>
<td>14.9</td>
<td>5.9</td>
<td>14.4</td>
</tr>
<tr>
<td>Sex (%Male)</td>
<td>47.8</td>
<td></td>
<td>52.0</td>
</tr>
<tr>
<td>Handedness (%Right)</td>
<td>93.2</td>
<td></td>
<td>92.9</td>
</tr>
</tbody>
</table>

MMSE Group Classifications per Tombaugh & McIntyre (1992)

<sup>a</sup>p = .006; <sup>b</sup>p < .0001
<table>
<thead>
<tr>
<th>Test</th>
<th>MMSE Level</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRS</td>
<td>18-23</td>
<td>67</td>
<td>-5.44</td>
<td>3.80</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>184</td>
<td>-2.42</td>
<td>2.26</td>
<td></td>
</tr>
<tr>
<td>BNT</td>
<td>18-23</td>
<td>90</td>
<td>-2.55</td>
<td>2.99</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>281</td>
<td>-1.11</td>
<td>1.96</td>
<td></td>
</tr>
<tr>
<td>COWA</td>
<td>18-23</td>
<td>90</td>
<td>-0.88</td>
<td>1.33</td>
<td>.019</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>279</td>
<td>-0.53</td>
<td>1.22</td>
<td></td>
</tr>
<tr>
<td>AVF</td>
<td>18-23</td>
<td>89</td>
<td>-1.24</td>
<td>1.18</td>
<td>.035</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>279</td>
<td>-0.91</td>
<td>1.30</td>
<td></td>
</tr>
<tr>
<td>Trails A</td>
<td>18-23</td>
<td>87</td>
<td>-1.30</td>
<td>2.59</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>279</td>
<td>-0.48</td>
<td>1.66</td>
<td></td>
</tr>
<tr>
<td>Trails B</td>
<td>18-23</td>
<td>73</td>
<td>-0.78</td>
<td>1.25</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>267</td>
<td>-0.72</td>
<td>1.74</td>
<td></td>
</tr>
</tbody>
</table>
# The Basic 2x2 Table

## Condition of Interest

<table>
<thead>
<tr>
<th></th>
<th>RCD +</th>
<th>RCD -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Negative</td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

**Factor (event)**

<table>
<thead>
<tr>
<th></th>
<th>MMSE 18-23</th>
<th>MMSE &gt; 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>False</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>Test</td>
<td>MMSE Level</td>
<td>% RCD</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>DRS</td>
<td>18-23</td>
<td>91.0</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>59.2</td>
</tr>
<tr>
<td>BNT</td>
<td>18-23</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>28.5</td>
</tr>
<tr>
<td>COWA</td>
<td>18-23</td>
<td>26.7</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>16.5</td>
</tr>
<tr>
<td>AVF</td>
<td>18-23</td>
<td>38.2</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>29.2</td>
</tr>
<tr>
<td>Trails A</td>
<td>18-23</td>
<td>36.8</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>16.5</td>
</tr>
<tr>
<td>Trails B</td>
<td>18-23</td>
<td>17.8</td>
</tr>
<tr>
<td></td>
<td>24-30</td>
<td>25.1</td>
</tr>
</tbody>
</table>
Conclusions

1. “Change” as a neurocognitive variable has its own unique statistical properties that are dependent on the reliability and stability of the tools used to assess change.

2. By taking into account the factors that affect change scores, it is possible to empirically identify changes that are relatively rare and likely to represent meaningful changes (outcomes) in a person’s cognitive status.

3. It is possible to use simple summary data in test manuals and research reports to create prediction equations to evaluate the significance between an individual’s observed and predicted retest scores, assisting the clinician to use research data obtained on populations to inform clinical decisions about individuals.
Selected References:


