NIH Toolbox Cognition Battery Normative Standards across Administration Platforms

Sensitivity and Validity in HIV infection

Kaitlin B. Casaletto, PhD
Assistant Professor
University of California, San Francisco
Thank you!!!

Coauthors:

Anya Umlauf, M.S.
Ian Abramson, PhD
Donald Franklin, BA
Bob Heaton, PhD
History of the NIH Toolbox-Cognitive Battery: Platform Development

2006-2012: NIH Toolbox Development and Normative Study (desktop-based)

2012: Finalized Toolbox (web-based) released to research community

2015: Revised Cognitive Battery normative standards released (based on desktop version)

2016: iPad-based Toolbox released

OUR QUESTION:
Can the normative standards be applied across administration platforms?
To Norm, or not to Norm?

Quantify *change* (i.e., “impairment”) cross-sectionally

Clinic

   Individual level

Research

   No or non-demographically matched control group
   Remove effects of demographics (reduce variance, covariates)
Development of Normative Standards

Demographically Corrected Normative Standards for the English Version of the NIH Toolbox Cognition Battery

Demographically Corrected Normative Standards for the Spanish Language Version of the NIH Toolbox Cognition Battery

6,859 NIHTB Normative Study

Test, confirm, validate...

Smoothing to achieve variance
Extract residuals for adjusted T-scores

Adults (18-85yo)
Children (3-17yo)

Race/Ethnicity (White/Asian, Black, Hispanic)
Race/Ethnicity (White/Asian, Black, Hispanic, Multiracial)
Importance of Test Invariance for Normative Standards

Change to what goes into a test, change what comes out of that test!

“Never change standardized test stimuli.” – V. Malcarne, PhD

If a new score ≠ old score, norms won’t work

Example: DCCS T-score = 50 + 10 * \([[(\text{DCCS scaled score}) - (12.92 - 9.84*(\text{age/100}) + 1.45*(\text{edu/10}) + 0.30*\text{male})] / [1.78 - 0.63*(\text{age/100}) + 0.16*(\text{edu/10}) + 0.32*\text{male}]} / 1.28\)
Aims

1. Determine comparability of NIH Toolbox-CB in detecting HIV-associated neurocognitive disorders (HAND) across platforms

2. Validate the NIH Toolbox-CB in a known neurocognitive disease, HIV-associated neurocognitive disorders (HAND)
Why HIV-associated neurocognitive disorders (HAND)?

~30-50% of HIV+ individuals

Mild-to-moderate fronto-striatal pattern

Weaknesses: processing speed, executive functions, memory retrieval

Strengths: language processing

Heterogeneous

Limited floor/ceiling effects

Pragmatics... HNRP
**Participants:**
n=83 iPad and n=530 web-based NIHTB-CB assessments

<table>
<thead>
<tr>
<th></th>
<th>iPad</th>
<th>Web</th>
<th><strong>Cohen’s d</strong></th>
<th>iPad</th>
<th>Web</th>
<th><strong>Cohen’s d</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HIV+</td>
<td>HIV-</td>
<td></td>
<td>HIV+</td>
<td>HIV-</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>57</td>
<td>26</td>
<td></td>
<td>404</td>
<td>126</td>
<td></td>
</tr>
<tr>
<td>Age, y</td>
<td>52.3 (13.7)</td>
<td>52.3 (13.3)</td>
<td>0.001</td>
<td>50.6 (12.3)</td>
<td>48.5 (15.5)</td>
<td>0.16</td>
</tr>
<tr>
<td>Edu, y</td>
<td>14.0 (2.6)</td>
<td>14.7 (2.9)</td>
<td>0.26</td>
<td>13.9 (2.5)</td>
<td>14.2 (2.4)</td>
<td>0.12</td>
</tr>
<tr>
<td>Sex, % M (n)</td>
<td>86% (49)</td>
<td>69.2% (18)</td>
<td>0.43</td>
<td>88.6% (358)</td>
<td>61.1% (77)</td>
<td>0.75</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>61.4% (35)</td>
<td>69.2% (18)</td>
<td>0.02</td>
<td>61.1% (247)</td>
<td>55.6% (70)</td>
<td>0.12</td>
</tr>
<tr>
<td>Black</td>
<td>21.2% (12)</td>
<td>30.8% (8)</td>
<td></td>
<td>21.5% (87)</td>
<td>26.2% (33)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>17.5%</td>
<td>0%</td>
<td></td>
<td>17.3% (70)</td>
<td>18.3% (23)</td>
<td></td>
</tr>
<tr>
<td>WRAT-4 Reading</td>
<td>102.1 (10.9)</td>
<td>104 (13.5)</td>
<td>0.16</td>
<td>101.0 (14.0)</td>
<td>105.0 (13.7)</td>
<td>0.29*</td>
</tr>
<tr>
<td>“Gold-standard” Impaired, %</td>
<td>43.9% (25)</td>
<td>26.9% (7)</td>
<td>0.35</td>
<td>47.5% (192)</td>
<td>14.3% (18)</td>
<td>0.71</td>
</tr>
</tbody>
</table>

**Note:** *p<0.05; “gold-standard” = HNRP comprehensive neuropsychological battery
Equivalency study: new individuals "matched" to those in Norming Study and administered both web and iPad.

**Web conversion**

1. Conversion formula extracted

2. Conversion formula applied to HNRP

**iPad conversion**

1. Conversion formula extracted

2. Conversion formula applied to HNRP
Equivalency smoothing usually helps...

N=158 iPad vs. Web matched on age (±5yrs), edu (±2yrs), sex, race/ethnicity, HIV serostatus

Cohen’s d: iPad > Web

- DCCS
- Flanker
- List Sort
- Pic Seq Memory
- Pattern Comp
- Vocab
- Reading

- Raw
- Equiv. Smooth (raw)
- T-score (demo adj.)
Norm-adjusted NIHTB-CB Fluid Scores are sensitive to HIV-infection, across platforms

Cohen’s d: HIV+ < HIV-

* Web-based (n=530) | iPad-based (n=83)

<table>
<thead>
<tr>
<th>Task</th>
<th>Cohen’s d</th>
<th>Web-based</th>
<th>iPad-based</th>
<th>Stat. Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCCS</td>
<td>0.4</td>
<td>0.3</td>
<td>0.5</td>
<td>*</td>
</tr>
<tr>
<td>Flanker</td>
<td>0.5</td>
<td>0.4</td>
<td>0.6</td>
<td>*</td>
</tr>
<tr>
<td>List Sort</td>
<td>0.3</td>
<td>0.2</td>
<td>0.4</td>
<td>*</td>
</tr>
<tr>
<td>Pic Seq</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td>*</td>
</tr>
<tr>
<td>Memory Comp</td>
<td>0.3</td>
<td>0.2</td>
<td>0.4</td>
<td>*</td>
</tr>
<tr>
<td>Fluid</td>
<td>0.5</td>
<td>0.4</td>
<td>0.6</td>
<td>*</td>
</tr>
<tr>
<td>Reading</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td>*</td>
</tr>
<tr>
<td>Vocab</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Crystal</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td></td>
</tr>
</tbody>
</table>

*p<0.05
NIHTB-CB is sensitive to “gold standard” HAND

Web and iPad, combined

All HIV+: no HAND > HAND

*all bars p’s<0.05
WRAT-4 and NIHTB-CB Crystallized scores

HIV+: $r=0.83^{**}$
HIV-: $r=0.82^{**}$
NIHTB-CB is sensitive to HAND

Normative standards can be statistically adjusted...
  ...not ideal...
  work in progress

NIHTB-CB Fluid scores can detect HIV+ difference
  Show effect with gold standard, comprehensive battery

NIHTB-CB Crystallized scores track with gold standard (WRAT4)
  Consider Crystallized based cut-offs to define Fluid “impairment”
Lessons learned

1. Tests need to be normed with the platforms/programs that will be used going forward!

2. Programs or Tests should never be changed after norming

3. If/when technology substantially advances/changes, re-norming may be necessary
Questions?