Assessing the Content Validity of Performance Outcome (PerfO) Measures

SEVENTH ANNUAL
PATIENT-REPORTED OUTCOME (PRO) CONSORTIUM WORKSHOP

April 27 - 28, 2016 ■ Silver Spring, MD
The views and opinions expressed in the following slides are those of the individual presenters and should not be attributed to their respective organizations/companies, the U.S. Food and Drug Administration, the Critical Path Institute, the PRO Consortium, or the ePRO Consortium.

These slides are the intellectual property of the individual presenters and are protected under the copyright laws of the United States of America and other countries. Used by permission. All rights reserved. All trademarks are the property of their respective owners.
Session Objectives

• To outline types of performance outcome (PerfO) assessments

• To discuss approaches to evaluating the content validity of PerfO assessments
Session Outline

• Introduction

• Case Study: Physical

• Case Study: Cognitive

• FDA Response and Comments

• Questions and Discussion
Session Participants

Moderator

– Elizabeth (Nicki) Bush, MHS – Research Scientist, Eli Lilly and Company

Presenters

– Rachel Ballinger, PhD – Lead Outcomes Researcher, Clinical Outcome Assessment, ICON Clinical Research
– Richard S.E. Keefe, PhD – Professor of Psychiatry, Psychology, and Neuroscience, Duke University Medical Center and CEO, NeuroCog Trials, Inc.

Panelists

– Michelle Campbell, PhD – Reviewer and Scientific Coordinator, COA Qualification Program, COA Staff, OND, CDER, FDA
– Stephen Joel Coons, PhD – Executive Director, Patient-Reported Outcome Consortium, Critical Path Institute
– Billy Dunn, MD – Director, Division of Neurology Products, OND, CDER, FDA
Introduction

Elizabeth (Nicki) Bush, MHS
Research Scientist, Eli Lilly and Company
What is a PerfO assessment?

- A clinical outcome assessment (COA)
- Measurement based on a task(s) performed by a patient according to instructions that is administered by a health care professional. Performance outcomes require patient cooperation and motivation. These include measures of gait speed (e.g., timed 25 foot walk test), memory recall, or other cognitive testing (e.g., digit symbol substitution test)\(^1\)

What is a PerfO assessment?

• A clinical outcome assessment (COA)
• Measurement based on a task(s) performed by a patient according to instructions that is administered by a health care professional. Performance outcomes require patient cooperation and motivation. These include measures of gait speed (e.g., timed 25 foot walk test), memory recall, or other cognitive testing (e.g., digit symbol substitution test).¹

• NOT a ClinRO

What is content validity?

• **Content validity** — Evidence from qualitative research demonstrating that the instrument measures the concept of interest including evidence that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use. Testing other measurement properties will not replace or rectify problems with content validity.

*FDA, Final PRO Guidance, 2009*
What is content validity?

• **Content validity** — Evidence from qualitative research demonstrating that the instrument measures the concept of interest including evidence that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use. Testing other measurement properties will not replace or rectify problems with content validity.

  *FDA, Final PRO Guidance, 2009*
PerfO assessments are similar to other COAs

• The foundations of content validity have not changed
  – Identify the concept(s) of interest. Are we measuring the right thing?
  – Assure the COA is appropriate. Are we measuring the concept in a way that is most relevant?
  – Assure the COA is comprehensive. Are we measuring the appropriate/core aspects of the concept?

• Interpretability is crucial
Is the PRO Guidance applicable to PerfO assessments?

• Mostly, yes:
  – Patient involvement
  – Iterative process
  – Measurement properties
  – Context of use
  – Interpretability

• But not always directly applicable:
  – Recall period
  – Response options
How might the PRO Guidance be applied to PerfO assessments?

• Specifically
  – Context of use (COU)
  – Concept of interest (COI)
  – Content Validity
Qualification of CLINICAL OUTCOME ASSESSMENTS (COAs)

V. Modify Instrument
- Identify a new COU
- Change wording of items, response options, recall period, or mode/method of administration/data collection
- Translate and culturally adapt
- Evaluate modifications using spokes I - IV
- Document all changes

Consider submitting to FDA for qualification of new COA, as appropriate.

IV. Longitudinal Evaluation of Measurement Properties/Interpretation Methods
- Assess ability to detect change and construct validity
- Identify responder definition(s)
- Provide guidelines for interpretation of treatment benefit and relationship to claim
- Document all results
- Update user manual

Submit to FDA for COA qualification as effectiveness endpoint to support claims.

III. Cross-sectional Evaluation of Other Measurement Properties
- Assess score reliability (test-retest or inter-rater) and construct validity
- Establish administration procedures & training materials
- Document measure development
- Prepare user manual

Consider submitting to FDA for COA qualification for use in exploratory studies prior to longitudinal evaluation.

I. Identify Context of Use (COU) and Concept of Interest (COI)
- Outline hypothesized concepts and potential claims
- Determine intended population
- Determine intended application/characteristics (type of scores, mode and frequency of administration)
- Perform literature/expert review
- Develop hypothesized conceptual framework
- Position COA within a preliminary endpoint model
- Document COU and COI

II. Draft Instrument and Evaluate Content Validity
- Obtain patient or other reporter input
- Generate new items
- Select recall period, response options and format
- Select mode/method of administration/data collection
- Conduct cognitive interviewing
- Pilot test draft instrument
- Finalize instrument content, format and scoring rule
- Document content validity
Spoke I – COU and COI

I. Identify Context of Use (COU) and Concept of Interest (COI)

- Outline hypothesized concepts and potential claims
- Determine intended population
- Determine intended application/characteristics (type of scores, mode and frequency of administration)
- Perform literature/expert review
- Develop hypothesized conceptual framework
- Position COA within a preliminary endpoint model
- Document COU and COI

- No differences between PerfO assessment and other COAs
- Hypothesized conceptual framework may have different headings
Hypothesized Conceptual Framework

Day-to-day functioning/Functional living skills

- Performance of Task 1
- Performance of Task 2
- Performance of Task 3
- Performance of Task 4
Concept elicitation is as crucial for PerfO assessments as other COAs
Here lie the main differences between PerfO assessments and other COAs in context of PRO Guidance. Not all are directly applicable, BUT think about the *spirit* of these steps...
Here lie the main differences between PerfO assessments and other COAs in context of PRO Guidance. Not all are directly applicable, BUT think about the *spirit* of these steps...

- **Language and concepts are meaningful and understandable**
- **Leads to meaningful, reliable, and interpretable score**
- **Consistent, reliable administration**
II. Draft Instrument and Evaluate Content Validity

- Obtain patient or other reporter input
- Generate new items
- Select recall period, response options and format
- Select mode/method of administration/data collection
- Conduct cognitive interviewing
- Pilot test draft instrument
- Finalize instrument content, format and scoring rule
- Document content validity
Spoke II – Content Validity

Consider understandability and relevance to day-to-day life; methods may not include traditional cognitive interviewing.

Pay special attention to uniformity of assessment administration; instructions to patients may affect motivation and compliance with the test.

Consider how each aspect affects uniformity.

II. Draft Instrument and Evaluate Content Validity

- Obtain patient or other reporter input
- Generate new items
- Select recall period, response options and format
- Select mode/method of administration/data collection
- Conduct cognitive interviewing
- Pilot test draft instrument
- Finalize instrument content, format and scoring rule
- Document content validity
II. Draft Instrument and Evaluate Content Validity

- Obtain patient or other reporter input
- Generate new items
- Select recall period, response options and format
- Select mode/method of administration/data collection
- Conduct cognitive interviewing
- Pilot test draft instrument
- Finalize instrument content, format and scoring rule
- Document content validity
Orthopedic Case Study

Rachel Ballinger, PhD
Lead Outcomes Researcher, Clinical Outcome Assessment, ICON Clinical Research
Study background

• Eli Lilly and Company program to evaluate compound for use in elective total hip replacement (eTHR), elective total knee replacement (eTKR) and hip fracture (HF) patients

• Need to validate select PerfO measures used in trials with intended trial population
  – Main psychometric PerfO validation study at multiple sites: Reliability (inter-rater, test-retest), Construct Validity (known-groups, convergent/divergent), Ability to Detect Change, Minimal Important Difference (MID) and Responder Definitions.
  – Substudy to assess content validation with sample of participants

• Approach per the standards of the FDA guidance for patient reported outcomes (2009) to the extent that these could be applied to PerfOs
## Study Design: Main Study

### Main study: longitudinal design with 3 visits

<table>
<thead>
<tr>
<th>Hip Fracture: all post-surgery (aged 65+, N:75)</th>
<th>Elective Knee/ Hip Replacement: two pre-, one- post surgery visits (aged 50+, N:200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timed up and go</td>
<td>Timed up and go</td>
</tr>
<tr>
<td>4-step stair climb</td>
<td>4-step stair climb</td>
</tr>
<tr>
<td>Repeated chair stands (x2)</td>
<td>Long stair climb</td>
</tr>
</tbody>
</table>

### Sub study: single qualitative telephone interview following visit
Study Design: Content Validation

• Study protocol and interview guide informed by
  – Methods outlined in FDA PRO Guidance
  – FDA Feedback

• The FDA was specifically interested to know how well they [the patients] believe the tests reflect their ability to function on a day-to-day basis, how the level of difficulty reflects the challenges they face in their daily function, and related topics.
Challenges

• Interview guide
  – Need to reflect issues of concern

• Availability of participants
  – Timeframe of interviews in relation to recent site visit

• Diminishing pool of potential participants
  – Reflect characteristics of sample in wider main study

• Saturation
Saturation Assessment

• Analysis codes
• Developed a data saturation summary grid
• Summaries developed and reviewed
• Per participant summary, for each PerfO:
  – Relevance, speed, and difficulty of the test = 9 summaries per participant
  – eTHR 72 summaries, HF 162 summaries
• Per participant ‘new element’
  – Between each summary and within each of the 3 themes the participant summary was compared to prior summaries to identify the new element(s) from each interview
  – eTHR 72 comparison summaries, HF 162 comparison summaries
• Overall Summary of New elements per theme and per PerfO = n:9
## Example: 4SC- overall speed (HF)

<table>
<thead>
<tr>
<th>Participant summary</th>
<th>ID#15</th>
<th>ID#16</th>
</tr>
</thead>
<tbody>
<tr>
<td>The participant said he did the steps at his normal speed without trying to go especially faster.</td>
<td></td>
<td>The participant said her norm is to move quite quickly and be slightly 'aggressive' when climbing stairs and she had no problem doing this in the test.</td>
</tr>
<tr>
<td>New element</td>
<td>[no new aspect: similar to no.5]</td>
<td>No problem with speed as her norm is to climb quickly/aggressively.</td>
</tr>
</tbody>
</table>
Key Results and Conclusions

• Main Study
  – Data from 75 HF, 98 eTHR and 103 eTKR patients at baseline
  – PerfOS suitable for use in eTHR, eTKR and HF patients
• Content Validation Study (sub-study)
  – Data from 8 eTHR patients and 18 HF patients
  – All HF and most eTHR participants related PerfOS to similar activities performed in daily life (albeit with some variation in specific aspects)
  – Most eTHR did not undertake longer stair climbs in daily life; some reported LSC gave them confidence for this in everyday life
  – All reported PerfOS to be relevant with a similar level of difficulty to daily activities
  – Participants generally reported finding each of the PerfOS easier to perform over time (across their visits), and the majority believed they would still see improvement as they continued to recover
Key Learnings

• PerfOs are unique
• A standardised approach is key
• Participants can distinguish between increased familiarity with PerfO and functional improvement
• PerfOs can impact patients’ confidence to perform certain activities
Acknowledgements

• All study participants and recruiting sites

• Lilly - Nicki Bush, April Naegeli, Olivier Benichou, MJay Shoenfeld, Elisa Gomez

• ICON - Helen Doll, Chloe Patel, Brittany Gentile, Magdi Vanya

• Former Oxford Outcomes / ICON -
  Cicely Kerr
  Annabel Nixon
  Paul Swinburn
  Sarah Hearn
  Katie Breheny
  Sarah Shingler
  Fiona Mowbray

Thank you!
Development and Validation of a Computerized Virtual Reality-based Assessment of Functional Capacity

Richard S.E. Keefe, PhD
Professor of Psychiatry, Psychology, and Neuroscience, Duke University Medical Center and CEO, NeuroCog Trials, Inc.
Disclosure

• The copyright for the Virtual Reality Functional Capacity Assessment Tool (VRFCAT) is held by my company NeuroCog Trials, Inc.
Definition of MCI Associated with AD

• “Persons with MCI commonly have mild problems performing complex functional tasks which they used to perform previously, such as paying bills, preparing a meal, or shopping.”

• “Lower performance in one or more cognitive domains, including memory, executive function, attention, language, and visuospatial skills.”

• Other aspects of cognition affected, such as working memory, information processing

Content of VRFCAT based upon activities that most challenge MCI patients

• **Functional capacity** refers to an “individual’s capacity for performing key tasks of daily living” (such as meal preparation or taking public transportation) as measured in a simulated clinic environment through completion of real world activities.”1 (Green et al. 2008; McKibbin et al. 2004; Bellack et al. 1994)

• The VRFCAT clearly addresses some of the activities that are of concern to patients with MCI, AD, and their caregivers
  1. Telephone use1, 2 (e.g., dialing numbers, answering phone, looking up numbers)
  2. Shopping1-3 (e.g., making purchases)
  3. Preparing meals1-3 (e.g., planning, preparing, and serving meals)
  4. Household chores1-3 (e.g., laundry, dishwashing, bed making)
  5. Transportation2 (e.g., using public transportation, driving a car)
  6. Responsibility for own medications2 (e.g., taking correct doses at scheduled times)
  7. Finances1, 2 (e.g., budgeting, writing checks, paying bills)
VRFCAT creates a realistic, interactive, and immersive environment consisting of 4 mini scenarios:

1. Planning a Meal in the Kitchen
2. Choosing and Paying for Bus to Grocery Store
3. Shopping and Purchasing Food in a Grocery Store
4. Choosing and Paying for Bus Home

The VRFCAT content is appropriate because it is related to what most people with MCI and schizophrenia struggle with in real life and it includes the core cognitive impairments of these disorders as determined by content experts.
# VRFCAT Scenarios and Objectives

## Mini Scenario Objectives 1-12

<table>
<thead>
<tr>
<th>Mini Scenario</th>
<th>Objectives 1-12</th>
<th>Cognitive Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apartment</td>
<td>1. Pick up the recipe on the counter</td>
<td>Visuospatial ability</td>
</tr>
<tr>
<td></td>
<td>2. Search for ingredients in your cabinets and refrigerator</td>
<td>Visuospatial ability</td>
</tr>
<tr>
<td></td>
<td>3. Access your recipe and cross off the ingredients that you already have in</td>
<td>Verbal and Visual Memory, Working Memory</td>
</tr>
<tr>
<td></td>
<td>your apartment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Pick up the billfold on the counter</td>
<td>Visuospatial ability</td>
</tr>
<tr>
<td></td>
<td>5. Exit the apartment and head to the bus stop (Game Element)</td>
<td></td>
</tr>
<tr>
<td>Bus to Store</td>
<td>6. Wait for the correct bus to the grocery store and then board it when it</td>
<td>Attention, Verbal Memory, Executive Functioning</td>
</tr>
<tr>
<td></td>
<td>arrives</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7. Add up the exact amount of bus fare in your hand and pay for the bus</td>
<td>Working Memory</td>
</tr>
<tr>
<td>Store</td>
<td>8. Select a food aisle to begin shopping</td>
<td>Executive Functioning</td>
</tr>
<tr>
<td></td>
<td>9. Continue shopping for the necessary food ingredients, and when finished</td>
<td>Attention, Visuospatial ability, Visual Memory,</td>
</tr>
<tr>
<td></td>
<td>check out</td>
<td>Verbal Memory, Executive Functioning</td>
</tr>
<tr>
<td></td>
<td>10. Add up the exact amount for your purchase and pay for groceries</td>
<td>Working Memory</td>
</tr>
<tr>
<td>Bus to Apartment</td>
<td>11. Wait for the correct bus to your apartment and then board it when it</td>
<td>Attention, Verbal Memory, Executive Functioning</td>
</tr>
<tr>
<td></td>
<td>arrives</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12. Add up the exact amount of bus fare in your hand and pay for the bus</td>
<td>Working Memory</td>
</tr>
</tbody>
</table>
• Collected extensive data on the experience of patients and other test-takers with regard to the instrument and how they understood the task, the goals of the task, how they interacted with the task elements, and of course their performance.

• Vast differences in the development of PerfO assessments and other COAs. Patients with cognitive impairment might not report accurately on their understanding of the PerfO in the cognitive interview.

• PerfO assessments have performance metrics to inform you whether someone understood the elements of the measure.
Scenario Versions (Alternate Forms)

Scenarios vary by:

1. Recipe and Ingredients
2. Ingredients in kitchen
3. Bus Fares
4. Monetary Amounts in Billfold
5. Purchase Amounts at Checkout

Scenarios are structurally and sequentially the same across versions
## Validation in Schizophrenia

*Indicates significant differences between HC and SZ at P < 0.05.*

<table>
<thead>
<tr>
<th></th>
<th>HC (N = 165)</th>
<th>SZ (N = 158)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, Mean (St Dev)</td>
<td>42.6 (13.93)</td>
<td>43.6 (11.84)</td>
</tr>
<tr>
<td>Male, N (%)</td>
<td>88 (53)</td>
<td>87 (55)</td>
</tr>
<tr>
<td>Non Hispanic, N (%)</td>
<td>136 (82)</td>
<td>128 (81)</td>
</tr>
<tr>
<td>English as Primary Language, N (%)</td>
<td>157 (95)</td>
<td>151 (96)</td>
</tr>
<tr>
<td>Unemployed, N (%)*</td>
<td>54 (33)</td>
<td>135 (85)</td>
</tr>
<tr>
<td>Comfortable with PC, N (%)*</td>
<td>160 (97)</td>
<td>140 (89)</td>
</tr>
<tr>
<td>Years of Education, Mean (St Dev)*</td>
<td>14.7 (2.41)</td>
<td>12.8 (1.99)</td>
</tr>
<tr>
<td>Mother’s Years of Education, Mean (St Dev)</td>
<td>12.9 (2.98)</td>
<td>12.5 (3.33)</td>
</tr>
</tbody>
</table>
Patients with Schizophrenia performed worse on all of the objectives.

Validation in Schizophrenia: Average Time to Complete each VRFCAT Objective
### Validation: Test-Retest and Practice Effects

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Visit 1 Mean (SD)</th>
<th>Visit 2 Mean (SD)</th>
<th>Cohen’s d</th>
<th>Intraclass Correlation Coefficient (ICC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HC</td>
<td>SZ</td>
<td>HC</td>
<td>SZ</td>
</tr>
<tr>
<td>VRFCAT Total Time T-score</td>
<td>50.1 (11.12)</td>
<td>32.3 (16.78)</td>
<td>50.9 (11.52)</td>
<td>31.8 (17.62)</td>
</tr>
<tr>
<td>VRFCAT Total Errors T-score</td>
<td>49.7 (11.48)</td>
<td>37.1 (22.74)</td>
<td>49.8 (12.94)</td>
<td>36.7 (22.07)</td>
</tr>
<tr>
<td>VRFCAT Progression T-score</td>
<td>49.8 (10.20)</td>
<td>40.4 (13.66)</td>
<td>50.3 (10.51)</td>
<td>40.8 (13.58)</td>
</tr>
<tr>
<td>UPSA-2-VIM*</td>
<td>83.4 (9.06)</td>
<td>70.7 (11.83)</td>
<td>86.7 (9.07)</td>
<td>74.5 (12.07)</td>
</tr>
</tbody>
</table>

*Indicates significant differences between Visit 1 and Visit 2 for both HC and SZ groups (p < 0.001).
### Pearson Correlation Coefficients between VRFCAT, UPSA-2-VIM & MCCB

<table>
<thead>
<tr>
<th>Assessment</th>
<th>VRFCAT Total Time T-score</th>
<th>VRFCAT Total Errors T-score</th>
<th>VRFCAT Progression T-score</th>
<th>MCCB</th>
</tr>
</thead>
<tbody>
<tr>
<td>VRFCAT Total Time T-score</td>
<td>---</td>
<td>0.75</td>
<td>0.60</td>
<td>0.68</td>
</tr>
<tr>
<td>VRFCAT Total Errors T-score</td>
<td>0.69</td>
<td>---</td>
<td>0.70</td>
<td>0.50</td>
</tr>
<tr>
<td>VRFCAT Progression T-score</td>
<td>0.70</td>
<td>0.64</td>
<td>---</td>
<td>0.35</td>
</tr>
<tr>
<td>MCCB Composite T-score</td>
<td>0.57</td>
<td>0.39</td>
<td>0.45</td>
<td>---</td>
</tr>
</tbody>
</table>

All correlations p-values were < 0.001
## Correlations of Real World Functioning with UPSA and VRFCAT

### Specific Levels of Functioning (SLOF)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPSA-VIM</td>
<td>.25**</td>
</tr>
<tr>
<td>VRFCAT Total Time</td>
<td>.22**</td>
</tr>
<tr>
<td>VRFCAT Total Errors</td>
<td>.29***</td>
</tr>
<tr>
<td>VRFCAT Progression</td>
<td>.17*</td>
</tr>
</tbody>
</table>

* p<.05, **p<.01, ***p<.001

N=158

UPSA-VIM, UCSD Performance-based Skills Assessment, Validation of Intermediate Measures version

VRFCAT, Virtual Reality Functional Capacity Assessment Tool
### Sample Demographics

<table>
<thead>
<tr>
<th></th>
<th>YA 18-30 yo (N=44)</th>
<th>OA 55-70 yo (N=41)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, Mean (St Dev)</td>
<td>25.8 (3.47)</td>
<td>60.8 (4.38)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Male, N (%)</td>
<td>24 (55)</td>
<td>17 (41)</td>
<td>0.224</td>
</tr>
<tr>
<td>Caucasian, N (%)</td>
<td>25 (57)</td>
<td>23 (56)</td>
<td>0.947</td>
</tr>
<tr>
<td>Years of Education, Mean (St Dev)</td>
<td>14.8 (2.28)</td>
<td>14.9 (2.95)</td>
<td>0.873</td>
</tr>
<tr>
<td>Employed, N (%)</td>
<td>30 (68)</td>
<td>12 (29)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Comfortable with PC, N (%)</td>
<td>44 (100)</td>
<td>37 (90)</td>
<td>0.035</td>
</tr>
</tbody>
</table>

### Cognitive Interview Results

<table>
<thead>
<tr>
<th></th>
<th>YA (N=44)</th>
<th>OA (N=41)</th>
<th>p-value¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleasantness, Mean (St Dev)</td>
<td>5.7 (1.47)</td>
<td>5.9 (1.36)</td>
<td>0.501</td>
</tr>
<tr>
<td>Ease of Use, Mean (St Dev)</td>
<td>6.8 (0.64)</td>
<td>6.1 (1.53)</td>
<td>0.004</td>
</tr>
<tr>
<td>Instructions, Mean (St Dev)</td>
<td>6.8 (0.70)</td>
<td>6.2 (1.41)</td>
<td>0.006</td>
</tr>
<tr>
<td>Realistic, Mean (St Dev)</td>
<td>6.0 (1.25)</td>
<td>6.1 (1.48)</td>
<td>0.468</td>
</tr>
</tbody>
</table>

Subject tolerability measures ranged from 1-7 with higher scores indicating higher levels of tolerability. P-values reflect Wilcoxon two sample rank sum analysis.

Atkins et al., J Prev Alz Dis 2015;2(2):121-127

YA = Younger Adults
OA = Older Adults
Evaluating Age Differences in Healthy Population

- Strong age-related differences in performance on total completion time, total errors, and total forced progressions \(p<.001\) for all

\[
\begin{align*}
\text{YA} &= \text{Younger Adults} \\
\text{OA} &= \text{Older Adults}
\end{align*}
\]

\cite{Atkins et al., 2014 (CTAD)}
Mean Completion Time on VRFCAT Objectives for Young and Older Adults

YA = Younger Adults
OA = Older Adults

- Cross off ingredients
- Shopping and paying For groceries
Validation: Evaluating Age Differences in Healthy Population

Mean Errors on VRFCAT Objectives for Young and Older Adults

YA = Younger Adults
OA = Older Adults

- Cross off ingredients
- Pay for bus
- Shop and pay for groceries
Validation: Evaluating Age Differences in Healthy Population

Correlations Between Other Functional Capacity Measures and Cognition

<table>
<thead>
<tr>
<th></th>
<th>MCCB Composite</th>
<th>TMT</th>
<th>BACSSC</th>
<th>HVLT</th>
<th>WMSIII</th>
<th>LNS</th>
<th>NAB</th>
<th>BVMT</th>
<th>Fluency</th>
<th>CPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>VRFCAT TIME T</td>
<td>0.66</td>
<td>0.48</td>
<td>0.47</td>
<td>0.49</td>
<td>0.47</td>
<td>0.62</td>
<td>0.50</td>
<td>0.47</td>
<td>0.35</td>
<td>0.45</td>
</tr>
<tr>
<td>VRFCAT Errors T</td>
<td>0.55</td>
<td>0.45</td>
<td>0.45</td>
<td>0.39</td>
<td>0.53</td>
<td>0.59</td>
<td>0.43</td>
<td>0.46</td>
<td>0.22</td>
<td>0.34</td>
</tr>
<tr>
<td>VRFCAT Progressions T</td>
<td>0.37</td>
<td>0.28</td>
<td>0.23</td>
<td>0.41</td>
<td>0.16</td>
<td>0.47</td>
<td>0.25</td>
<td>0.33</td>
<td>0.21</td>
<td>0.22</td>
</tr>
<tr>
<td>Bus T</td>
<td>0.41</td>
<td>0.09</td>
<td>0.12</td>
<td>0.41</td>
<td>0.34</td>
<td>0.48</td>
<td>0.28</td>
<td>0.32</td>
<td>0.44</td>
<td>0.29</td>
</tr>
<tr>
<td>Recipe T</td>
<td>0.05</td>
<td>-0.08</td>
<td>-0.07</td>
<td>0.24</td>
<td>-0.09</td>
<td>0.27</td>
<td>-0.06</td>
<td>0.03</td>
<td>0.13</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

**NOTE:** Uncorrected T-scores are used for the VRFCAT and MCCB measures
MCCB Subtests include: Trail Making Test, Part A (TMT); Brief Assessment of Cognition Symbol Coding (BACSSC); Hopkins Verbal Learning Test-Revised (HVLT); Wechsler Memory Scale-III (WMSIII); Letter Number Span (LNS); Neuropsychological Assessment Battery Mazes (NAB); Brief Visuospatial Memory Test – Revised (BVMT); Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT); Continuous Performance test-Identical Pairs (CPT).
Qualification of **CLINICAL OUTCOME ASSESSMENTS** (COAs)

**V. Modify Instrument**
- Identify a new COU
- Change wording of items, response options, recall period, or mode/method of administration/data collection
- Translate and culturally adapt
- Evaluate modifications using spokes I - IV
- Document all changes

*Consider submitting to FDA for qualification of new COA, as appropriate.*

**IV. Longitudinal Evaluation of Measurement Properties/Interpretation Methods**
- Assess ability to detect change and construct validity
- Identify responder definition(s)
- Provide guidelines for interpretation of treatment benefit and relationship to claim
- Document all results
- Update user manual

*Submit to FDA for COA qualification as effectiveness endpoint to support claims.*

**III. Cross-sectional Evaluation of Other Measurement Properties**
- Assess score reliability (test-retest or inter-rater) and construct validity
- Establish administration procedures & training materials
- Document measure development
- Prepare user manual

*Consider submitting to FDA for COA qualification for use in exploratory studies prior to longitudinal evaluation.*

**I. Identify Context of Use (COU) and Concept of Interest (COI)**
- Outline hypothesized concepts and potential claims
- Determine intended population
- Determine intended application/characteristics (type of scores, mode and frequency of administration)
- Perform literature/expert review
- Develop hypothesized conceptual framework
- Position COA within a preliminary endpoint model
- Document COU and COI

**II. Draft Instrument and Evaluate Content Validity**
- Obtain patient or other reporter input
- Generate new items
- Select recall period, response options and format
- Select mode/method of administration/data collection
- Conduct cognitive interviewing
- Pilot test draft instrument
- Finalize instrument content, format and scoring rule
- Document content validity

---

U.S. Food and Drug Administration
Center for Drug Evaluation and Research
Office of New Drugs
http://www.fda.gov/Drugs
Richard Keefe, Ph.D.
CEO, NeuroCog Trials
Professor of Psychiatry, Psychology and Neuroscience, Duke University Medical Center
Richard.Keefe@duke.edu
Assessing Performance Outcome Measures in Mild Cognitive Impairment due to Alzheimer’s Disease: A C-Path Case Study

J. Scott Andrews, PharmD
Research Scientist, Eli Lilly and Company
Measuring Function Across the Continuum of Alzheimer’s Disease

**Target Patient Population**

- **Pre-clinical AD**
- **MCI due to AD**
- **Mild AD**
- **Moderate AD**
- **Severe AD**

**Positive Amyloid PET Imaging Scan and/or Aβ CSF**

- **Asymptomatic**
  - Slight memory deficit
  - 26-30 MMSE

- **No Functional impairment**
  - Some difficulty with complex function (iADLs)

**Memory impairment**

- 20-26 MMSE

**Functional impairment**

- 10-20 MMSE

- 0-10 MMSE*

**Behavior Symptoms**

- Some Assistance
- Nursing home

*MMSE cutoffs representative*
Measuring Function Across the Continuum of Alzheimer’s Disease

**Target Patient Population**

- Pre-clinical AD
- MCI due to AD
- Mild AD
- Moderate AD
- Severe AD

**Positive Amyloid PET Imaging Scan and/or Aβ CSF**

- Asymptomatic
  - Slight memory deficit
    - 26-30 MMSE
  - No Functional impairment

- Memory impairment
  - 20-26 MMSE
  - Some difficulty with complex function (iADLs)

- 10-20 MMSE
  - Functional impairment

- 0-10 MMSE*
  - Behavior Symptoms
  - Some Assistance
  - Nursing home

*MMSE cutoffs representative
COA Selection in MCI due to AD

- PRO – Patient report not reliable
- OsbRo – Informant report lacks sensitivity
- ClinRo – Clinician report not appropriate
  - PerfO ?

Oppportunities
- Measurement properties
- Patient-focus
- Direct evidence?

Challenges
- Operational feasibility
- Cross-cultural applicability
- Content validity?
- Interpretation?
Pathway for Measurement Selection

Consistent Conceptual Framework

Instrumental ADL Task Performance

- Telephone use
- Shopping
- Preparing meals
- Household chores
- Transportation
- Responsibility for own medications
- Finances

Consensus Development Workshop

Instrument Review & Selection

Expert Opinion

Literature Review

Qualitative Research & Consultation
## PerfO Instrument Review

<table>
<thead>
<tr>
<th>Measure</th>
<th>Method</th>
<th>Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPSA</td>
<td>Role-play test</td>
<td>shopping/meal prep, communication, finances, transportation, planning</td>
</tr>
<tr>
<td>University of California San Diego Performance-based Skills Assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VRFCAT</td>
<td>Computer-based</td>
<td>transportation, finances, household management, planning</td>
</tr>
<tr>
<td>Virtual Reality Functional Capacity Assessment Tool</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Functional Capacity: Comparison of Informant (ADCS-ADL) versus Performance-Based (UPSA) Measure

![Bar chart comparing functional capacity scores between ADCS-ADL and UPSA measures for HC, MCI, and AD groups.](chart.png)

ADCS-ADL: AD Cooperative Study ADL Scale
Attendees:
• FDA Division of Neurology Products
• FDA Clinical Outcome Assessment
• FDA Office of Biostatistics
• Expert Consultants

Outcome: explore qualification of UPSA

- Concept of interest can be refined through qualification process. Encourage preliminary discussion.
  - FDA DNP, COA

- Gaps that should be addressed with current PerfO measures:
  • Content validity
  • Psychometrics
  • Interpretation
  • Learning effects
  - FDA DNP, COA & Biostatistics

Are PerfOs potentially appropriate for qualification as a co-primary measure? Yes.
  - FDA DNP

PerfOs can capture meaningful concepts and real-world translation shouldn’t be seen as an obstacle.
  - FDA DNP
FDA response and comments
Panel Discussion

Moderator
  – *Elizabeth (Nicki) Bush, MHS* – Research Scientist, Eli Lilly and Company

Presenters
  – *Rachel Ballinger, PhD* – Lead Outcomes Researcher, Clinical Outcome Assessment, ICON Clinical Research
  – *Richard S.E. Keefe, PhD* – Professor of Psychiatry, Psychology, and Neuroscience, Duke University Medical Center and CEO, NeuroCog Trials, Inc.

Panelists
  – *Michelle Campbell, PhD* – Reviewer and Scientific Coordinator, COA Qualification Program, COA Staff, OND, CDER, FDA
  – *Stephen Joel Coons, PhD* – Executive Director, Patient-Reported Outcome Consortium, Critical Path Institute
  – *Billy Dunn, MD* – Director, Division of Neurology Products, OND, CDER, FDA
Questions?
Backup slides
Q: What do you consider when choosing between types of COAs?

<table>
<thead>
<tr>
<th>Review Division</th>
<th>Disease/Condition</th>
<th>Indication and/or Claim(s) Description</th>
<th>Outcome of Interest</th>
<th>COA (COA Type)</th>
<th>COA Context of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>METABOLISM AND ENDOCRINOLOGY PRODUCTS</td>
<td>Muscle wasting disorder (lower extremity functional decline in patients with hip fracture)</td>
<td>To be determined</td>
<td>Lower-extremity functional decline</td>
<td>Usual Gait Speed (UGS) and the Short Physical Performance Battery Test (SPPB) (performance outcome)(^1)</td>
<td>Persons age 65 years and older who have diminished muscle mass and strength and decreased function that is a result of a hip fracture</td>
</tr>
<tr>
<td>METABOLISM AND ENDOCRINOLOGY PRODUCTS</td>
<td>Sarcopenia</td>
<td>To be determined</td>
<td>Physical functioning</td>
<td>Patient Reported Outcome Measurement System (PROMIS) – Physical Function item bank (patient-reported outcome)(^2)</td>
<td>Adult patients with sarcopenia</td>
</tr>
</tbody>
</table>

\(^1\) Submitter: Aging in Motion Coalition of the Alliance for Aging Research  
\(^2\) Submitter: PROMIS Network Center
### Some PerfO Assessments included in the FDA’s pilot COA Compendium

<table>
<thead>
<tr>
<th>Review Division</th>
<th>Disease/Condition</th>
<th>Indication and/or Claim(s) Description</th>
<th>Outcome of Interest</th>
<th>COA (COA Type)</th>
<th>COA Context of Use</th>
<th>COA Qualification Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TRANSPLANT AND OPHTHALMOLOGY PRODUCTS</strong></td>
<td>Neovascular (wet) age-related macular degeneration</td>
<td>Treatment of age-related macular degeneration</td>
<td>Best corrected visual acuity</td>
<td>Visual acuity (performance outcome)</td>
<td>Adult patients with age-related macular degeneration</td>
<td>Not applicable</td>
</tr>
<tr>
<td><strong>CARDIOVASCULAR AND RENAL PRODUCTS</strong></td>
<td>Chronic thromboembolic pulmonary hypertension (CTEPH)</td>
<td>Treatment of persistent/recurrent CTEPH after surgical treatment or inoperable CTEPH to improve exercise capacity and WHO functional class</td>
<td>Exercise capacity</td>
<td>6-Minute Walking Distance (performance outcome)</td>
<td>Adult patients with CTEPH</td>
<td>Not applicable</td>
</tr>
<tr>
<td><strong>CARDIOVASCULAR AND RENAL PRODUCTS</strong></td>
<td>Pulmonary arterial hypertension</td>
<td>Treatment of pulmonary arterial hypertension</td>
<td>Exercise capacity</td>
<td>6-minute Walking Distance (performance outcome)</td>
<td>Adult patients with pulmonary arterial hypertension</td>
<td>Not applicable</td>
</tr>
<tr>
<td><strong>NEUROLOGY PRODUCTS</strong></td>
<td>Alzheimer’s disease: Mild cognitive impairment due to Alzheimer’s disease (MCI due to AD)</td>
<td>To be determined</td>
<td>Day-to-day functioning (instrumental activities of daily living)</td>
<td>Currently unnamed (performance outcome tool to assess instrumental activities of daily living (IADLs))</td>
<td>Adults (&gt;45 years with mild cognitive impairment due to Alzheimer’s disease (MCI due to AD))</td>
<td>Submitter: Critical Path Institute: PRO Consortium’s Cognition Working Group</td>
</tr>
<tr>
<td>Review Division</td>
<td>Disease/Condition</td>
<td>Indication and/or Claim(s) Description</td>
<td>Outcome of Interest</td>
<td>COA (COA Type)</td>
<td>COA Context of Use</td>
<td>COA Qualification Information</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------</td>
<td>---------------------------------------</td>
<td>---------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>NEUROLOGY PRODUCTS</td>
<td>Multiple Sclerosis (MS)</td>
<td>To be determined</td>
<td>“MS disability” or simply “disability” characterized as neurological or neuropsychological deficits that result in limitation in activities, participation, or roles caused by MS that are understood to be important</td>
<td>New Clinical Outcome Assessment Instrument for Use in Clinical Trials of Medical Products to Treat Multiple Sclerosis (MS) (performance outcome)</td>
<td>Adults living with relapsing-remitting or progressive forms of MS</td>
<td>Submitter: Critical Path Institute Multiple Sclerosis Outcome Assessments Consortium (MSOCAC)</td>
</tr>
<tr>
<td>GASTROENTEROLOGY AND INBORN ERRORS PRODUCTS</td>
<td>Mucopolysaccharidosis I (MPS I) (Hurler and Hurler-Scheie forms of MPS I)</td>
<td>Improvement in walking capacity</td>
<td>Walking capacity</td>
<td>6-Minute Walk Test (performance outcome)</td>
<td>Pediatric and/or adult patients with MPS I</td>
<td></td>
</tr>
<tr>
<td>METABOLISM AND ENDOCRINOLOGY PRODUCTS</td>
<td>Muscle wasting disorder (lower extremity functional decline in patients with hip fracture)</td>
<td>To be determined</td>
<td>Lower-extremity functional decline</td>
<td>Usual Gait Speed (UGS) and the Short Physical Performance Battery Test (SPPB) (performance outcome)</td>
<td>Persons age 65 years and older who have diminished muscle mass and strength and decreased function that is a result of a hip fracture</td>
<td>Submitter: Aging in Motion Coalition of the Alliance for Aging Research</td>
</tr>
</tbody>
</table>
VRFCAT Future Updates

- Enhanced study and subject set-up
- Graphical Upgrade – Unreal Game Engine 4.0
- Audit Trail Functionality
- ‘Clinical Trial Mode’ – Locking settings, etc.
- Protections for data from deletion/editing, enhanced security
- Provide sites read-only access to data on server
- Required user credentials and Fingerprint Scanning for Subject Verification

'Clinical Trial Mode' – Locking settings, etc.

Provide sites read-only access to data on server

Audit Trail Functionality

Graphical Upgrade – Unreal Game Engine 4.0

Protections for data from deletion/editing, enhanced security

Required user credentials and Fingerprint Scanning for Subject Verification
Study Aims

The study assessed the validity, sensitivity and reliability of the VRFCAT in patients with schizophrenia (SZ) and healthy controls (HC), specifically:

- The discriminability of patients with schizophrenia and healthy controls
- Test-retest reliability
- Practice effects
- The relationship between VRFCAT outcomes and cognitive performance on the MATRICS Consensus Cognitive Battery (MCCB)
- Comparison between VRFCAT performance and UPSA-2-VIM
  - Note, the UPSA-2-VIM assesses the same five domains as the UPSA-1

Funded by NIMH SBIR phase 1 and phase 2 grants
Recruitment Methods

- 166 HCs and 158 patients with SZ were recruited from three sites:
  1. University of California San Diego
  2. University of Miami Miller School of Medicine
  3. University of South Carolina

NOTE: One HC was removed due to extremely low test scores (7 SD below mean)
Data Collection Methods and Analyses

- MCCB administered at Visit 1
- The VRFCAT and UPSA-2-VIM were completed at Visit 1 and 2, which was 7 to 14 days later
- Items on the VRFCAT were compared for the HCs and SZs

Analyses examined:
- Test-retest reliability
- Performance differences
- Correlations between VRFCAT measures, the MCCB Composite T-score and the UPSA-2-VIM Total Score
## Demographics

<table>
<thead>
<tr>
<th></th>
<th>Healthy Controls</th>
<th>Schizophrenia Patients</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Age</td>
<td>165</td>
<td>42.6</td>
<td>13.94</td>
<td>158</td>
</tr>
<tr>
<td>Years of Education</td>
<td>165</td>
<td>14.7</td>
<td>2.41</td>
<td>157</td>
</tr>
<tr>
<td>Mother’s Education</td>
<td>155</td>
<td>12.9</td>
<td>2.98</td>
<td>142</td>
</tr>
<tr>
<td>% (N)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>53 (88)</td>
<td></td>
<td>56 (88)</td>
<td>0.18</td>
</tr>
<tr>
<td>Unemployed</td>
<td>33 (54)</td>
<td></td>
<td>85 (135)</td>
<td>92.40</td>
</tr>
<tr>
<td>Comfortable with Computer</td>
<td>97 (160)</td>
<td></td>
<td>89 (140)</td>
<td>8.53</td>
</tr>
<tr>
<td>Hispanic</td>
<td>18 (29)</td>
<td></td>
<td>19 (30)</td>
<td>0.11</td>
</tr>
<tr>
<td>English Primary Language</td>
<td>95 (157)</td>
<td></td>
<td>96 (151)</td>
<td>0.03</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>56 (92)</td>
<td></td>
<td>47 (75)</td>
<td>3.33</td>
</tr>
<tr>
<td>African American</td>
<td>38 (63)</td>
<td></td>
<td>48 (76)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>6 (10)</td>
<td></td>
<td>4 (7)</td>
<td></td>
</tr>
</tbody>
</table>
Results: Visit 1 data for SZ group

The SCoRS, PANSS, and SLOF were only administered to the SZ group

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician SCoRS Total</td>
<td>38.2 (9.88)</td>
</tr>
<tr>
<td>PANSS Total</td>
<td>71.6 (21.93)</td>
</tr>
<tr>
<td>Clinician SLOF Total</td>
<td>120.8 (14.42)</td>
</tr>
</tbody>
</table>

- Schizophrenia Cognition Rating Scale (SCoRS)
- Positive and Negative Syndrome Scale (PANSS)
- Specific Level Of Functioning (SLOF) Total Scores
Results: Ability to Discriminate Between SZs and HCs

- The 3 VRFCAT summary measures, the MCCB Composite Score, and the UPSA-2-VIM all demonstrated significant differences between HC and SZ at the Visit 1.

<table>
<thead>
<tr>
<th></th>
<th>HC (N = 165)</th>
<th>SZ (N = 158)</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCCB Composite Score, Mean (SD)*</td>
<td>44.0 (13.19)</td>
<td>28.1 (12.91)</td>
<td>1.22</td>
</tr>
<tr>
<td>VRFCAT Total Time T-score, Mean (SD)*</td>
<td>49.7 (11.51)</td>
<td>32.5 (16.60)</td>
<td>1.21</td>
</tr>
<tr>
<td>VRFCAT Total Errors T-score, Mean (SD)*</td>
<td>49.4 (11.62)</td>
<td>37.6 (22.37)</td>
<td>0.67</td>
</tr>
<tr>
<td>VRFCAT Progression T-score, Mean (SD)*</td>
<td>49.7 (10.16)</td>
<td>40.5 (13.62)</td>
<td>0.77</td>
</tr>
<tr>
<td>UPSA-2-VIM, Mean (SD)*</td>
<td>83.2 (9.03)</td>
<td>71.0 (11.85)</td>
<td>1.16</td>
</tr>
</tbody>
</table>

* Indicates significant differences between HC and SZ at the 0.05 significance level.
Results: VRFCAT Total Time T-score Discrimination

[Histogram showing frequency distribution of VRFCAT Total Time T-score for HCs and SZs]
Results: Comparison of Performance on Different Measures at Visit 1

![Graph showing the comparison of VRFCAT Total Time T-score and UPSA-2-VIM with MCCB Composite Score.](image)
### Results: Evaluating Age Differences in Healthy Population

**Functional capacity and cognitive performance by age**

<table>
<thead>
<tr>
<th>VRFCAT Summary Measures</th>
<th>YA (N=44) Mean ± SD</th>
<th>OA (N=41) Mean ± SD</th>
<th>p-value</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Time (minutes)</td>
<td>11.8 ± 2.09</td>
<td>15.0 ± 3.28</td>
<td>&lt; 0.001</td>
<td>1.2</td>
</tr>
<tr>
<td>Total Errors</td>
<td>1.1 ± 1.50</td>
<td>3.1 ± 3.35</td>
<td>&lt; 0.001</td>
<td>0.8</td>
</tr>
<tr>
<td>Total Progressions</td>
<td>0.1 ± 0.21</td>
<td>0.5 ± 0.81</td>
<td>&lt; 0.001</td>
<td>0.9</td>
</tr>
<tr>
<td>Total Bus Schedule Checks</td>
<td>3.5 ± 1.90</td>
<td>3.7 ± 1.73</td>
<td>0.463</td>
<td>0.1</td>
</tr>
<tr>
<td>Total Recipe Checks</td>
<td>12.4 ± 4.81</td>
<td>11.8 ± 4.99</td>
<td>0.475</td>
<td>-0.1</td>
</tr>
<tr>
<td>UPSA-2-VIM Total Score</td>
<td>84.4 ± 8.63</td>
<td>83.1 ± 8.88</td>
<td>0.562</td>
<td>0.1</td>
</tr>
<tr>
<td>MCCB Composite T-Score (uncorrected)</td>
<td>49.0 ± 11.67</td>
<td>36.9 ± 12.00</td>
<td>&lt; 0.001</td>
<td>1.0</td>
</tr>
</tbody>
</table>

- Older subjects took an average of 3 minutes longer to complete the VRFCAT and made an average of 2 more errors during the test

YA = Younger Adults
OA = Older Adults
Results: Evaluating Age Differences in Healthy Population

- VRFCAT Total Time demonstrated good test-retest reliability (ICC=.80 in young adults; ICC=.64 in older adults) and non-significant practice effects
- VRFCAT Total Time was correlated with cognitive performance on MCCB (r=.79 in YA, r=.66 in OA)
Results from this study suggest the VRFCAT has:

- **Good test-retest** reliability in patients and healthy controls in different age groups
- **Strong correlations** with the MCCB and UPSA-2-VIM in patient populations and HCs in different age groups
- **Strong discrimination** between patients and healthy controls
- **Minimal practice effects** in all groups
- Almost no patients or controls at ceiling at baseline

These data provide support for the VRFCAT as a co-primary measure of functional capacity
SBIR Phase 2b Commercialization Plan.
Funding approved for May, 2016

• Aim 1: Establish normative data for the US/English VRFCAT
• Aim 2: Translation and software implementation of multicultural test content
• Aim 3: Conduct cognitive debriefing studies for multicultural VRFCAT versions in Russia, Poland, Italy, and Switzerland
• Aim 4: Validation of the multicultural VRFCAT versions in each country/culture
### Results: Evaluating Age Differences in Healthy Population

#### Practice Effects and Test-Retest Reliability for the VRFCAT and UPSA-2-VIM by Age Group

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Visit 1 Mean (SD)</th>
<th>Visit 2 Mean (SD)</th>
<th>Difference Mean (SD)</th>
<th>Cohen’s $d^1$</th>
<th>Intraclass Correlation Coefficient (ICC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YA</td>
<td>OA</td>
<td>YA</td>
<td>OA</td>
<td>YA</td>
</tr>
<tr>
<td>VRFCAT Total Time (minutes)</td>
<td>11.8 (2.10)</td>
<td>14.6 (2.52)</td>
<td>11.5 (2.25)</td>
<td>14.3 (3.45)</td>
<td>0.3 (1.38)</td>
</tr>
<tr>
<td>VRFCAT Total Errors</td>
<td>1.1 (1.46)</td>
<td>2.8 (3.04)</td>
<td>0.9 (1.28)</td>
<td>2.8 (4.65)</td>
<td>0.2 (1.43)</td>
</tr>
<tr>
<td>VRFCAT Total Progressions</td>
<td>0.0 (0.22)</td>
<td>0.5 (0.72)</td>
<td>0.0 (0.22)</td>
<td>0.4 (0.93)</td>
<td>0.0 (0.22)</td>
</tr>
<tr>
<td>Total Bus Schedule Checks</td>
<td>3.4 (1.91)</td>
<td>3.5 (1.67)</td>
<td>3.1 (1.70)</td>
<td>3.8 (2.39)</td>
<td>0.3 (2.00)</td>
</tr>
<tr>
<td>Total Recipe Checks</td>
<td>12.5 (4.91)</td>
<td>11.9 (5.10)</td>
<td>12.2 (5.46)</td>
<td>11.9 (5.23)</td>
<td>0.3 (4.50)</td>
</tr>
<tr>
<td>UPSA-2-VIM</td>
<td>84.8 (8.45)</td>
<td>83.4 (8.98)</td>
<td>87.6 (8.21)</td>
<td>86.2 (9.56)</td>
<td>2.7 (5.76)**</td>
</tr>
</tbody>
</table>

- Practice effects for all VRFCAT measures were small and insignificant in both age groups.
- A practice effect of 2.7 points ($d=0.3$) for the UPSA-2-VIM was observed in both older and younger adults ($p=0.018$ and $p=0.005$, respectively).