Assessing the Content Validity of Performance Outcome (PerfO) Measures

SEVENTH ANNUAL
PATIENT-REPORTED OUTCOME (PRO) CONSORTIUM WORKSHOP

April 27 - 28, 2016 ■ Silver Spring, MD

Disclaimer

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.

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.
- *J. Scott Andrews, PharmD* – Research Scientist, Eli Lilly and Company and Co-chair PRO Consortium's Cognition WG

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]

[1]FDA, COA Glossary of terms:
http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm370262.htm

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

[1]FDA, COA Glossary of terms: http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm370262.htm

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)

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

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.

IV. Longitudinal Evaluation of Measurement Properties/ Interpretation Methods

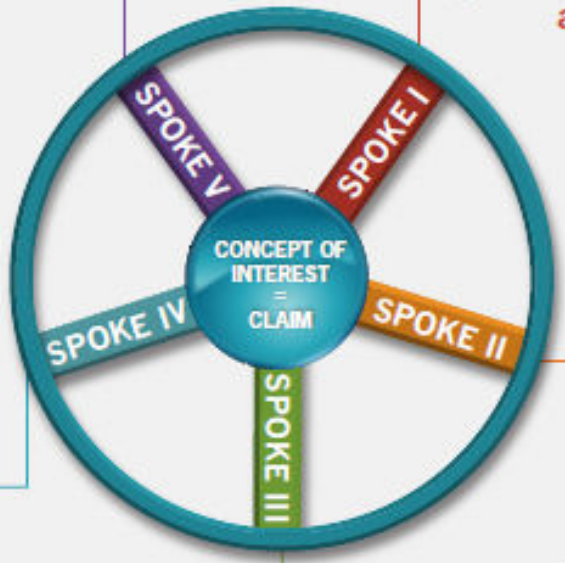
- Assess ability to detect change and construct validity
- Identify responder definition(s)
- Provide guidelines for interpretation of treatment benefit and relationship to claim
- Document all results
- Update user manual

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

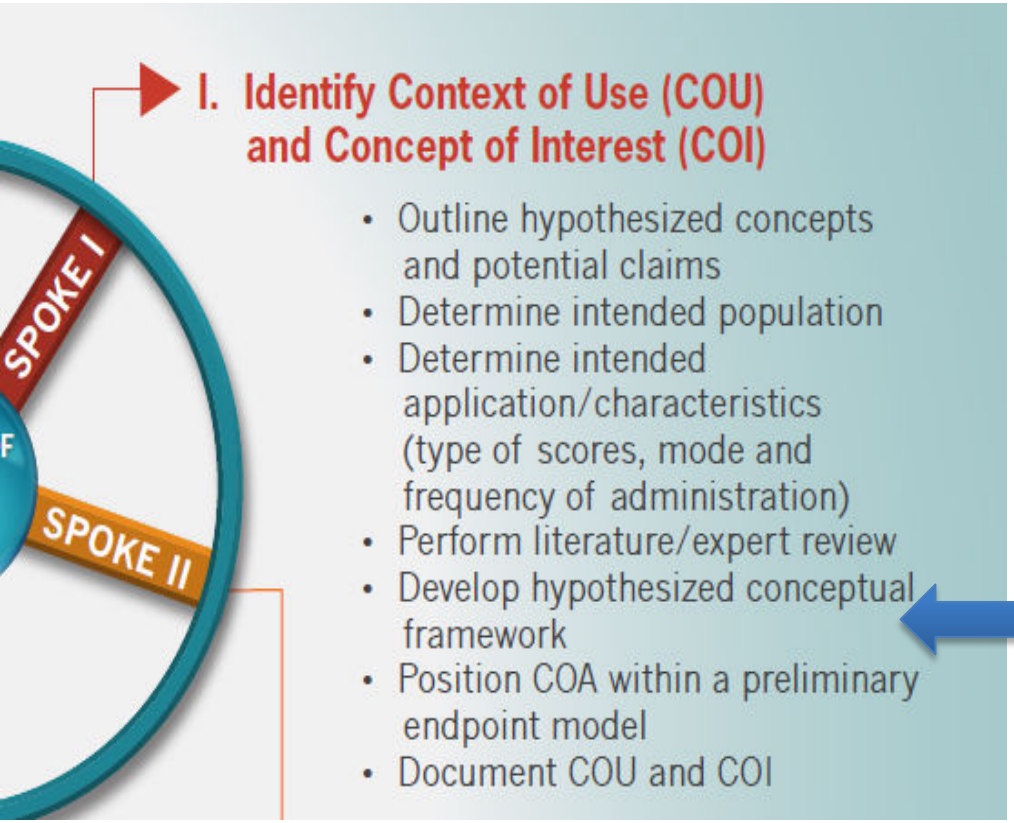
V. Modify Instrument

- 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.

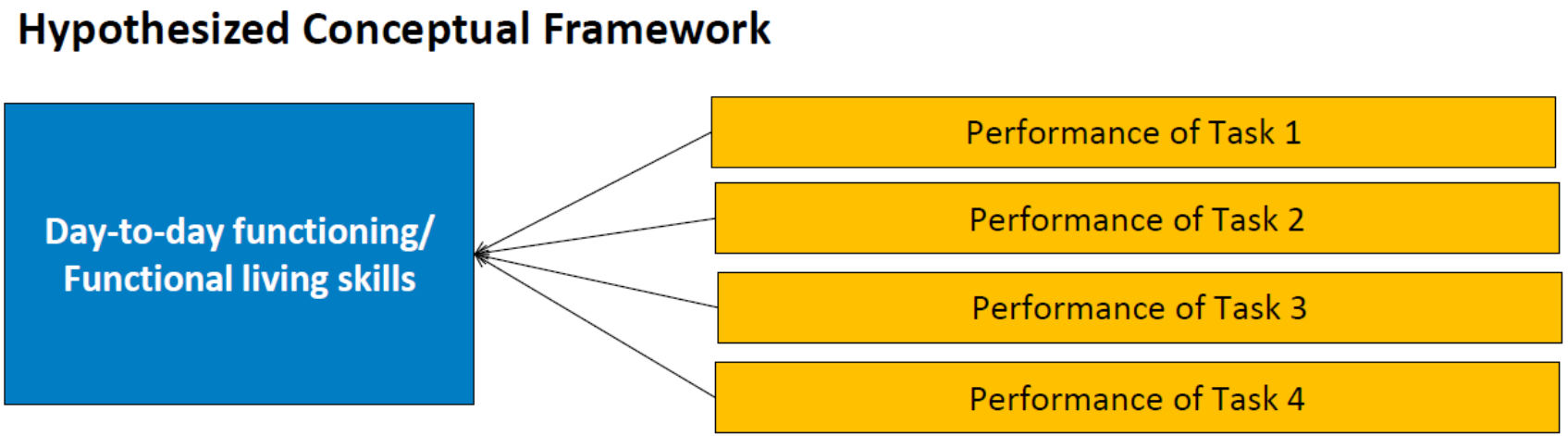


Spoke I – COU and COI



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

Conceptual framework example from Cognition Working Group



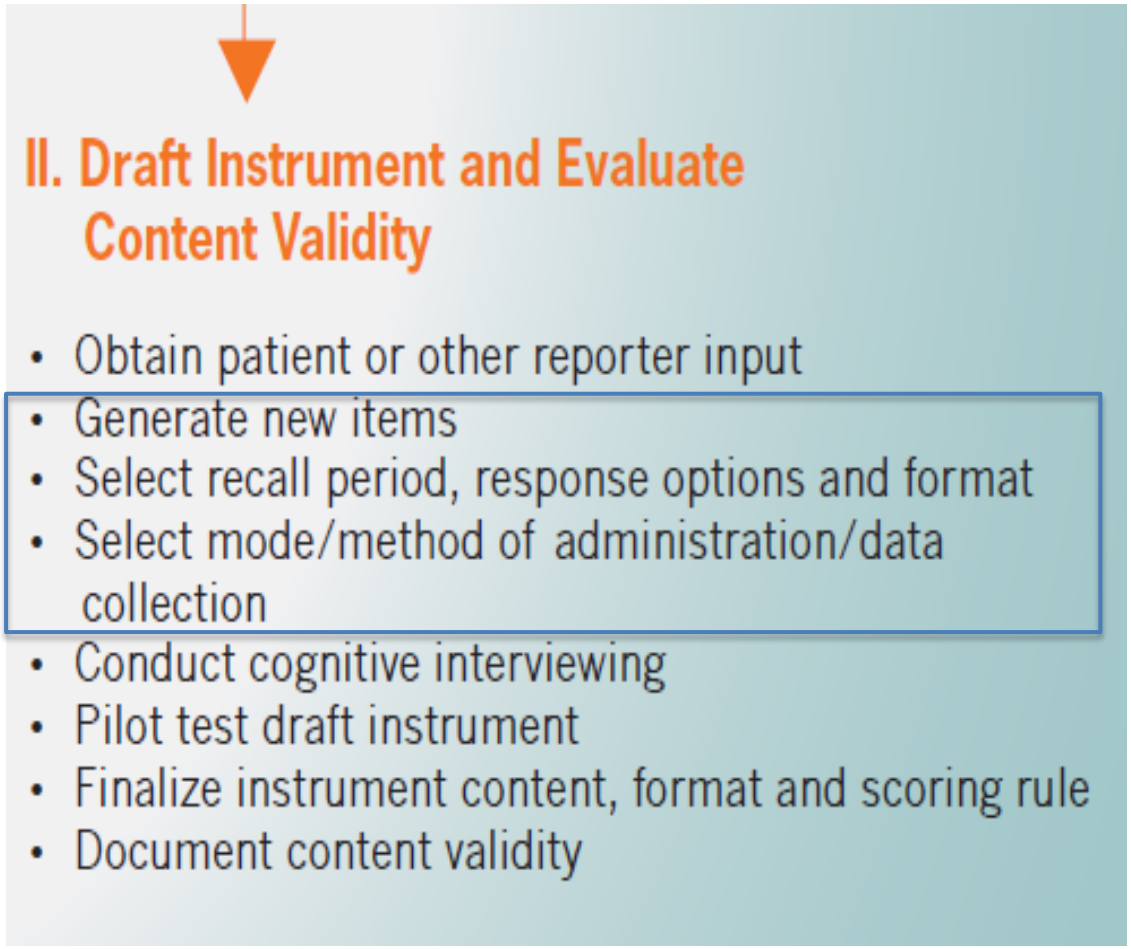
Spoke II – Content Validity

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

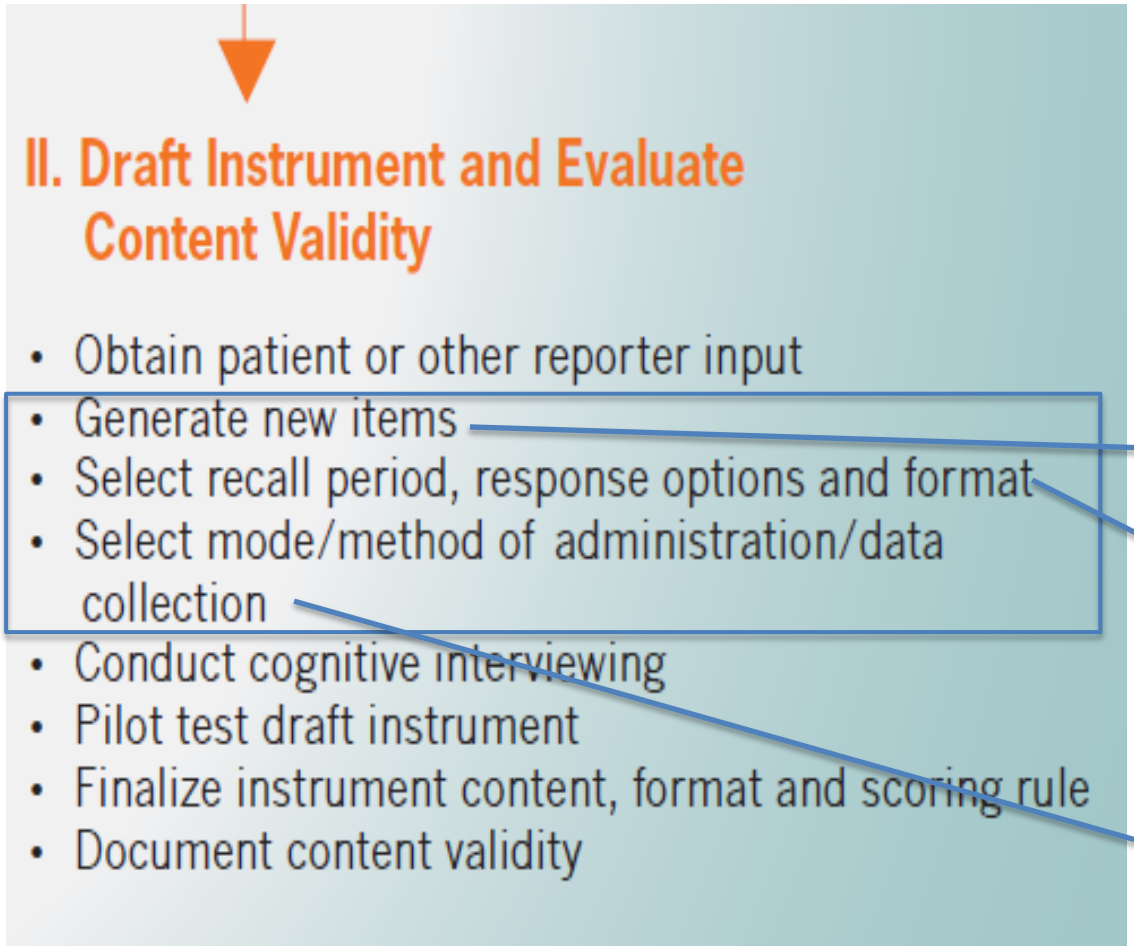
Concept elicitation is as crucial for PerfO assessments as other COAs

Spoke II – Content Validity



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…

Spoke II – Content Validity



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

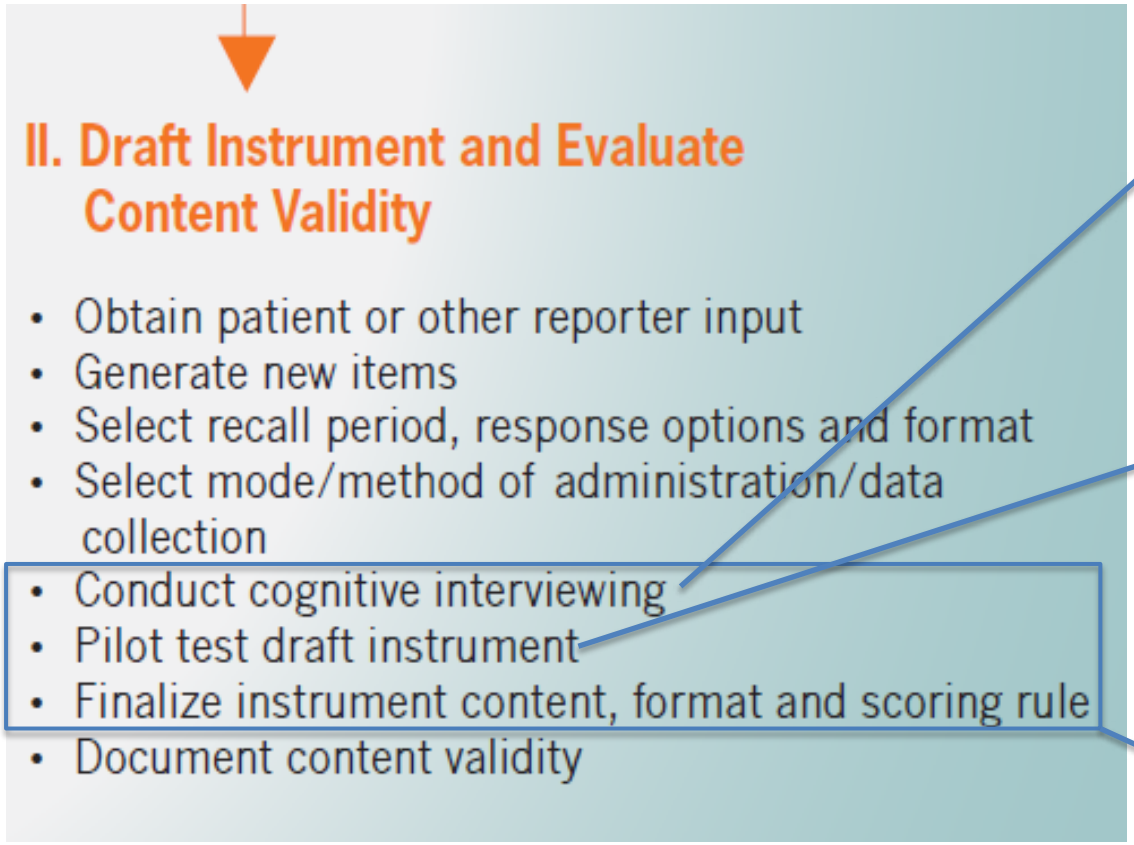
Consistent, reliable administration

Spoke II – 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

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

Spoke II – 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

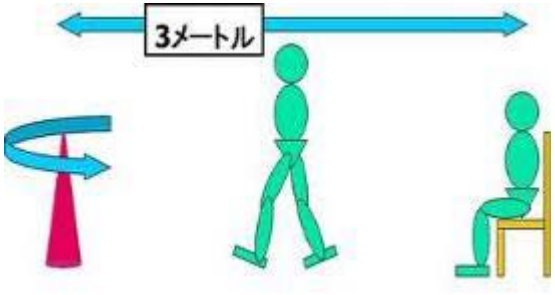
- Approach per the standards of the FDA guidance for patient reported outcomes (2009) to the extent that these could be applied to PerfOs

Study Design: Main Study

PRO CONSORTIUM
CRITICAL PATH INSTITUTE

Main study: longitudinal design with 3 visits

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



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)

	ID#15	ID#16
Participant summary	The participant said he did the steps at his normal speed without trying to go especially faster.	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.
New element	[no new aspect: similar to no.5]	No problem with speed as her norm is to climb quickly/aggressively.

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

Albert MS, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association*. 2011;7(3):270-279.

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 use[1, 2] (e.g., dialing numbers, answering phone, looking up numbers)
  2. Shopping[1-3] (e.g., making purchases)
  3. Preparing meals[1-3] (e.g., planning, preparing, and serving meals)
  4. Household chores[1-3] (e.g., laundry, dishwashing, bed making)
  5. Transportation[2] (e.g., using public transportation, driving a car)
  6. Responsibility for own medications[2] (e.g., taking correct doses at scheduled times)
  7. Finances[1, 2] (e.g., budgeting, writing checks, paying bills)

VRFCAT

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	Cognitive Domain
Apartment	1.	Pick up the recipe on the counter	Visuospatial ability
	2.	Search for ingredients in your cabinets and refrigerator	Visuospatial ability Executive Functioning
	3.	Access your recipe and cross off the ingredients that you already have in your apartment	Verbal and Visual Memory, Working Memory
	4.	Pick up the billfold on the counter	Visuospatial ability
	5.	Exit the apartment and head to the bus stop (Game Element)	
Bus to Store	6.	Wait for the correct bus to the grocery store and then board it when it arrives	Attention, Verbal Memory, Executive Functioning
	7.	Add up the exact amount of bus fare in your hand and pay for the bus	Working Memory
Store	8.	Select a food aisle to begin shopping	Executive Functioning
	9.	Continue shopping for the necessary food ingredients, and when finished check out	Attention, Visuospatial ability, Visual Memory Verbal Memory, Executive Functioning
	10.	Add up the exact amount for your purchase and pay for groceries	Working Memory
Bus to Apartment	11.	Wait for the correct bus to your apartment and then board it when it arrives	Attention, Verbal Memory, Executive Functioning
	12.	Add up the exact amount of bus fare in your hand and pay for the bus	Working Memory

VRFCAT Content Development

- 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.

VIDEO

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

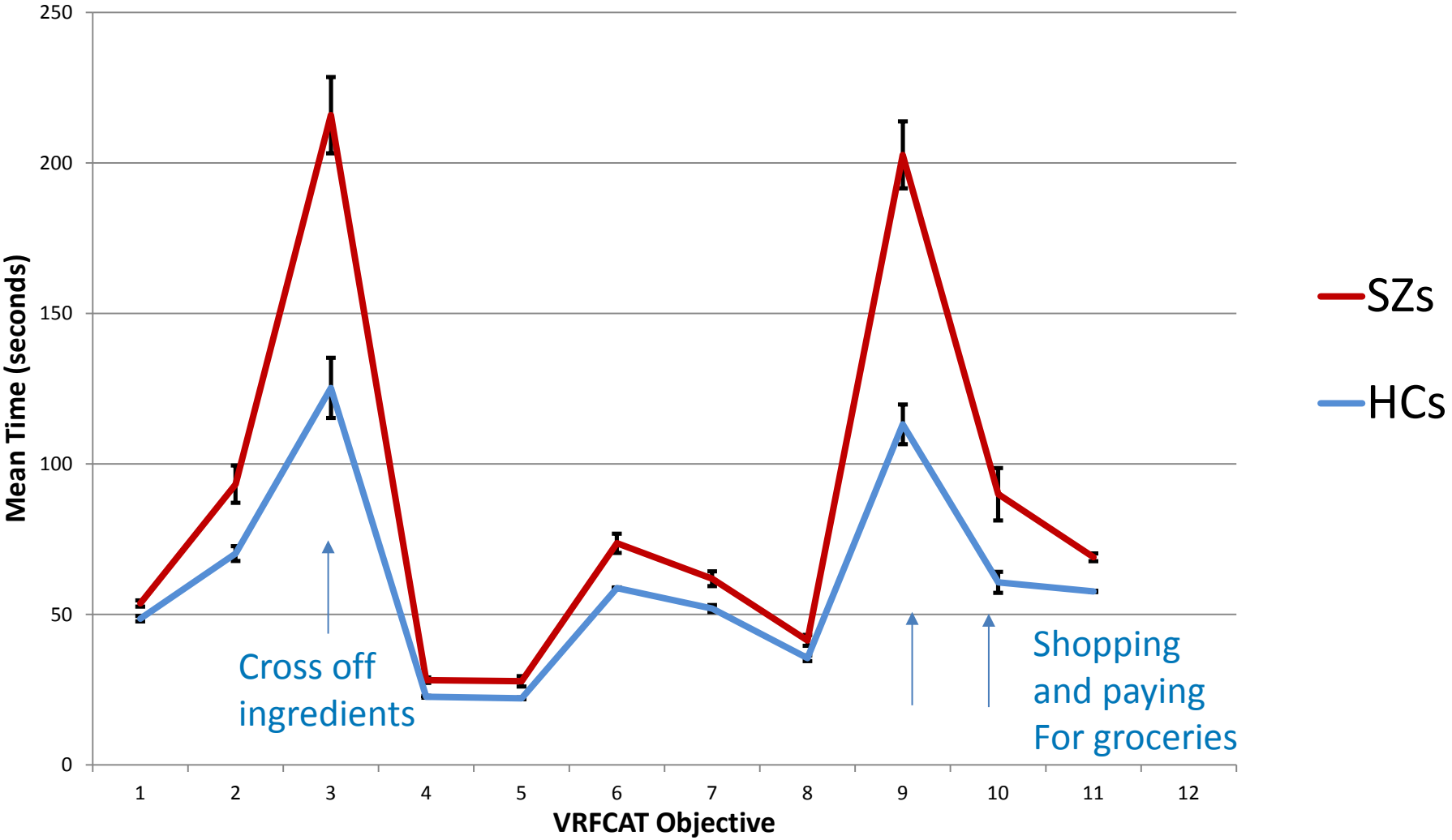
Validation in Schizophrenia

	HC (N = 165)	SZ (N = 158)
Age, Mean (St Dev)	42.6 (13.93)	43.6 (11.84)
Male, N (%)	88 (53)	87 (55)
Non Hispanic, N (%)	136 (82)	128 (81)
English as Primary Language, N (%)	157 (95)	151 (96)
Unemployed, N(%)*	54 (33)	135 (85)
Comfortable with PC, N (%)*	160 (97)	140 (89)
Years of Education, Mean (St Dev)*	14.7 (2.41)	12.8 (1.99)
Mother's Years of Education, Mean (St Dev)	12.9 (2.98)	12.5 (3.33)

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

Validation in Schizophrenia: Average Time to Complete each VRFCAT Objective

Patients with Schizophrenia performed worse on all of the objectives.



Validation: Test-Retest and Practice Effects

Assessment	Visit 1 Mean (SD)		Visit 2 Mean (SD)		Cohen's d		Intraclass Correlation Coefficient (ICC)	
	HC	SZ	HC	SZ	HC	SZ	HC	SZ
VRFCAT Total Time T-score	50.1 (11.12)	32.3 (16.78)	50.9 (11.52)	31.8 (17.62)	0.07	-0.03	0.65	0.81
VRFCAT Total Errors T-score	49.7 (11.48)	37.1 (22.74)	49.8 (12.94)	36.7 (22.07)	0.01	-0.02	0.54	0.65
VRFCAT Progression T-score	49.8 (10.20)	40.4 (13.66)	50.3 (10.51)	40.8 (13.58)	0.05	0.03	0.29	0.61
UPSA-2-VIM*	83.4 (9.06)	70.7 (11.83)	86.7 (9.07)	74.5 (12.07)	0.36	0.32	0.75	0.78

*Indicates significant differences between Visit 1 and Visit 2 for both HC and SZ groups ($p < 0.001$).

Validation: Relationship To Other Measures

Pearson Correlation Coefficients between VRFCAT, UPSA-2-VIM & MCCB

Healthy Controls Schizophrenia Patients

Assessment	VRFCAT Total Time T-score	VRFCAT Total Errors T-score	VRFCAT Progression T-score	MCCB
VRFCAT Total Time T-score	---	0.75	0.60	0.68
VRFCAT Total Errors T-score	0.69	---	0.70	0.50
VRFCAT Progression T-score	0.70	0.64	---	0.35
MCCB Composite T-score	0.57	0.39	0.45	---

All correlations p-values were < 0.001

Correlations of Real World Functioning with UPSA and VRFCAT

	Specific Levels of Functioning (SLOF)
UPSA-VIM	.25**
VRFCAT Total Time	.22**
VRFCAT Total Errors	.29***
VRFCAT Progression	.17*

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

Evaluating Age Differences in Healthy Population

PRO CONSORTIUM
CRITICAL PATH INSTITUTE

Sample Demographics

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

Cognitive Interview Results

	YA (N=44)	OA (N=41)	p-value[1]
Pleasantness, Mean (St Dev)	5.7 (1.47)	5.9 (1.36)	0.501
Ease of Use, Mean (St Dev)	6.8 (0.64)	6.1 (1.53)	0.004
Instructions, Mean (St Dev)	6.8 (0.70)	6.2 (1.41)	0.006
Realistic, Mean (St Dev)	6.0 (1.25)	6.1 (1.48)	0.468

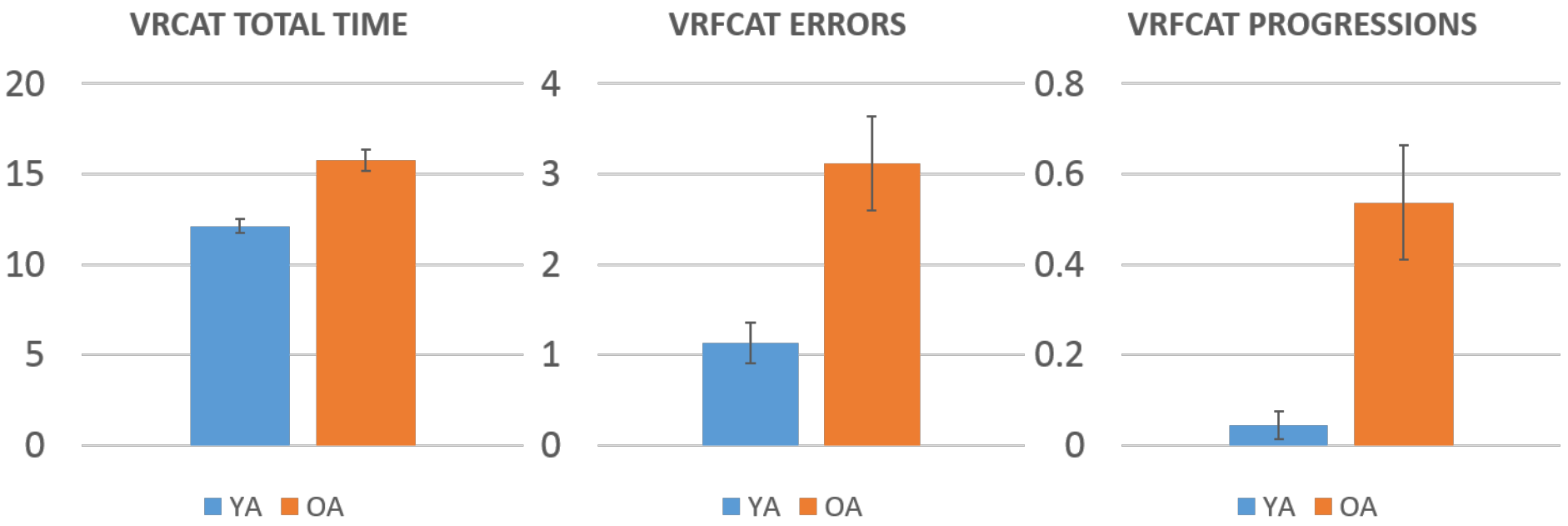
Subject tolerability measures ranged from 1-7 with higher scores indicating higher levels of tolerability.
P-values reflect Wilcoxon two sample rank sum analysis.

YA = Younger Adults
OA = Older Adults

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

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)



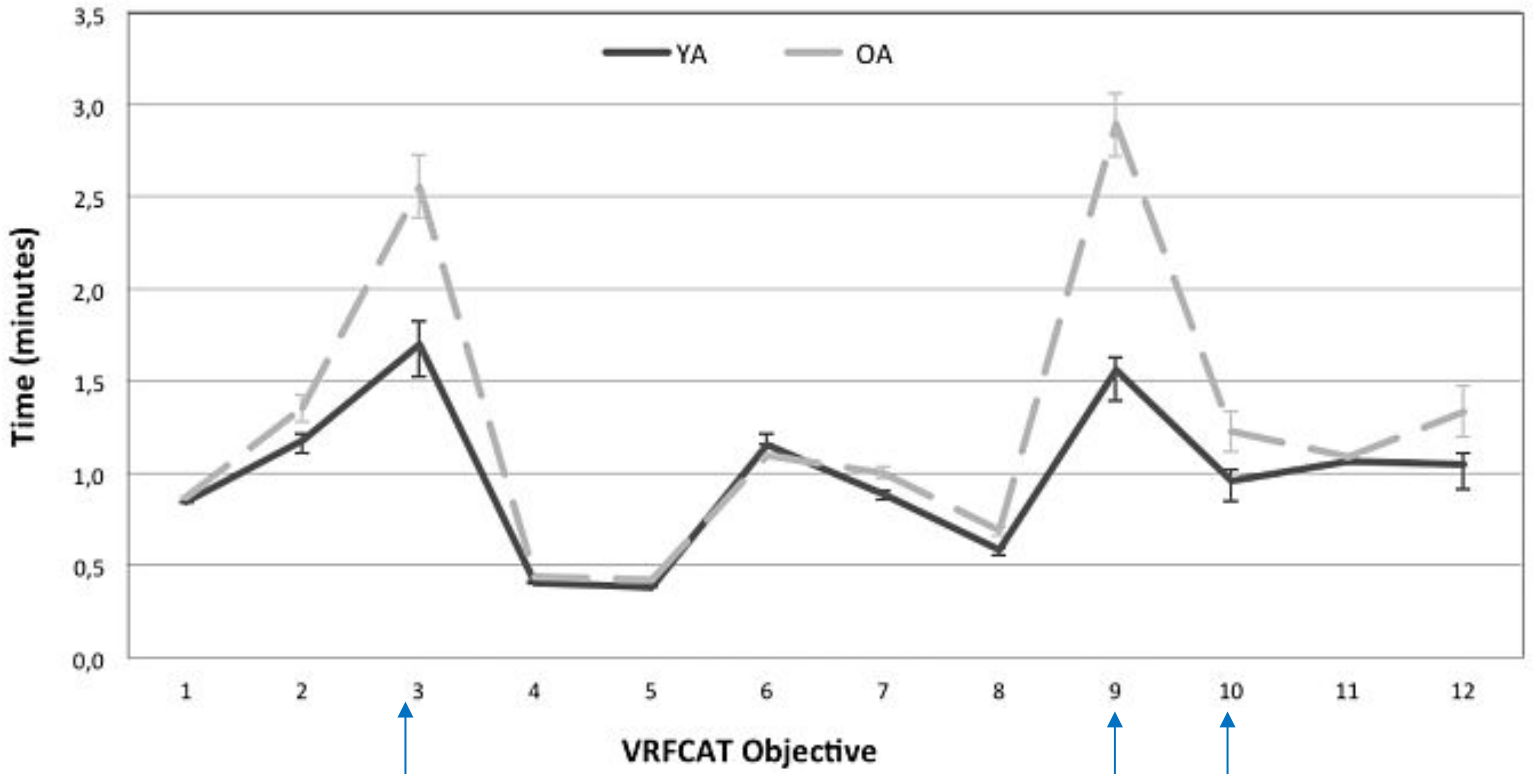
YA = Younger Adults
OA = Older Adults

Atkins et al., 2014 (CTAD)

Validation: Evaluating Age Differences in Healthy Population

PRO CONSORTIUM CRITICAL PATH INSTITUTE

Mean Completion Time on VRFCAT Objectives for Young and Older Adults

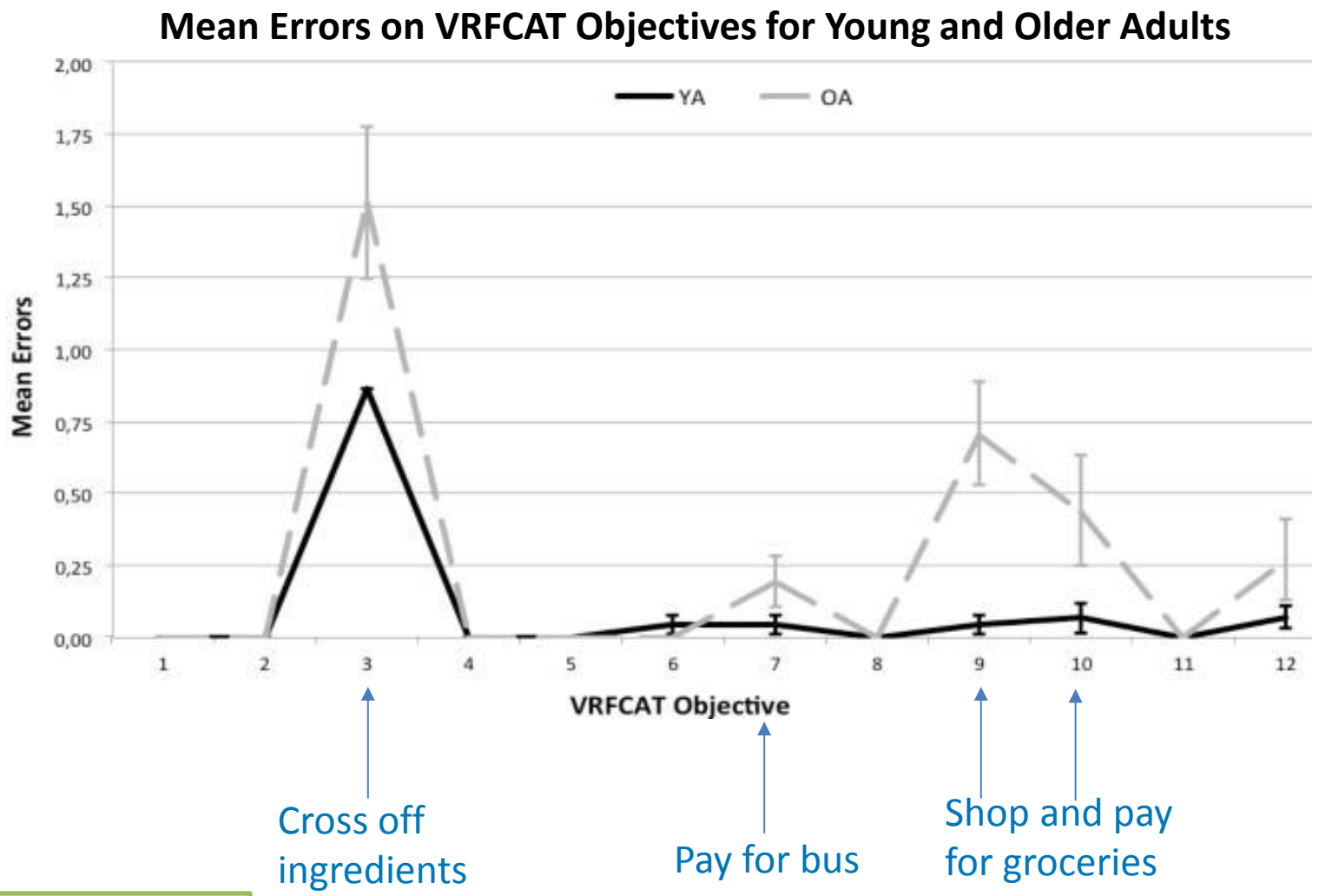


Cross off ingredients

Shopping and paying For groceries

YA = Younger Adults
OA = Older Adults

Validation: Evaluating Age Differences in Healthy Population



YA = Younger Adults
OA = Older Adults

Validation: Evaluating Age Differences in Healthy Population

PRO CONSORTIUM CRITICAL PATH INSTITUTE

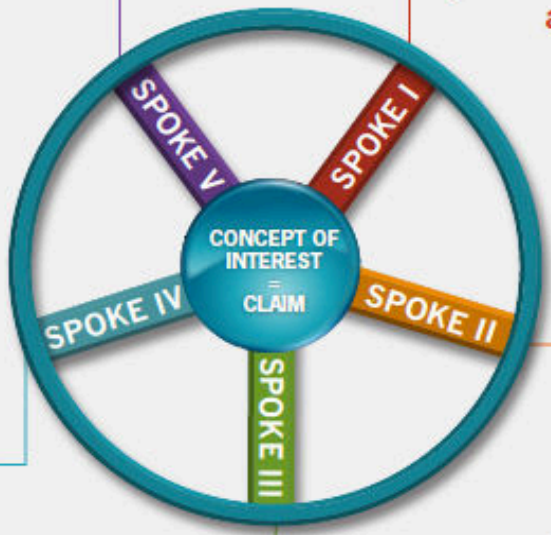
Correlations Between Other Functional Capacity Measures and Cognition

Older Adults	MCCB Composite	TMT	BACSSC	HVLT	WMSIII	LNS	NAB	BVMT	Fluency	CPT
		Speed of Processing	Speed of Processing	Verbal Learning	Working Memory	Working Memory	Reasoning and Problem Solving	Visual Learning	Speed of Processing	Attention /Vigilance
VRFCAT TIME T	0.66	0.48	0.47	0.49	0.47	0.62	0.50	0.47	0.35	0.45
VRFCAT Errors T	0.55	0.45	0.45	0.39	0.53	0.59	0.43	0.46	0.22	0.34
VRFCAT Progressions T	0.37	0.28	0.23	0.41	0.16	0.47	0.25	0.33	0.21	0.22
Bus T	0.41	0.09	0.12	0.41	0.34	0.48	0.28	0.32	0.44	0.29
Recipe T	0.05	-0.08	-0.07	0.24	-0.09	0.27	-0.06	0.03	0.13	-0.01

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)



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

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.

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.

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.

U.S. Food and Drug Administration
Center for Drug Evaluation and Research
Office of New Drugs
http://www.fda.gov/Drugs

For more information

Richard Keefe, Ph.D.
CEO, NeuroCog Trials
Professor of Psychiatry, Psychology and Neuroscience, Duke University Medical Center
Richard.Keefe@duke.edu

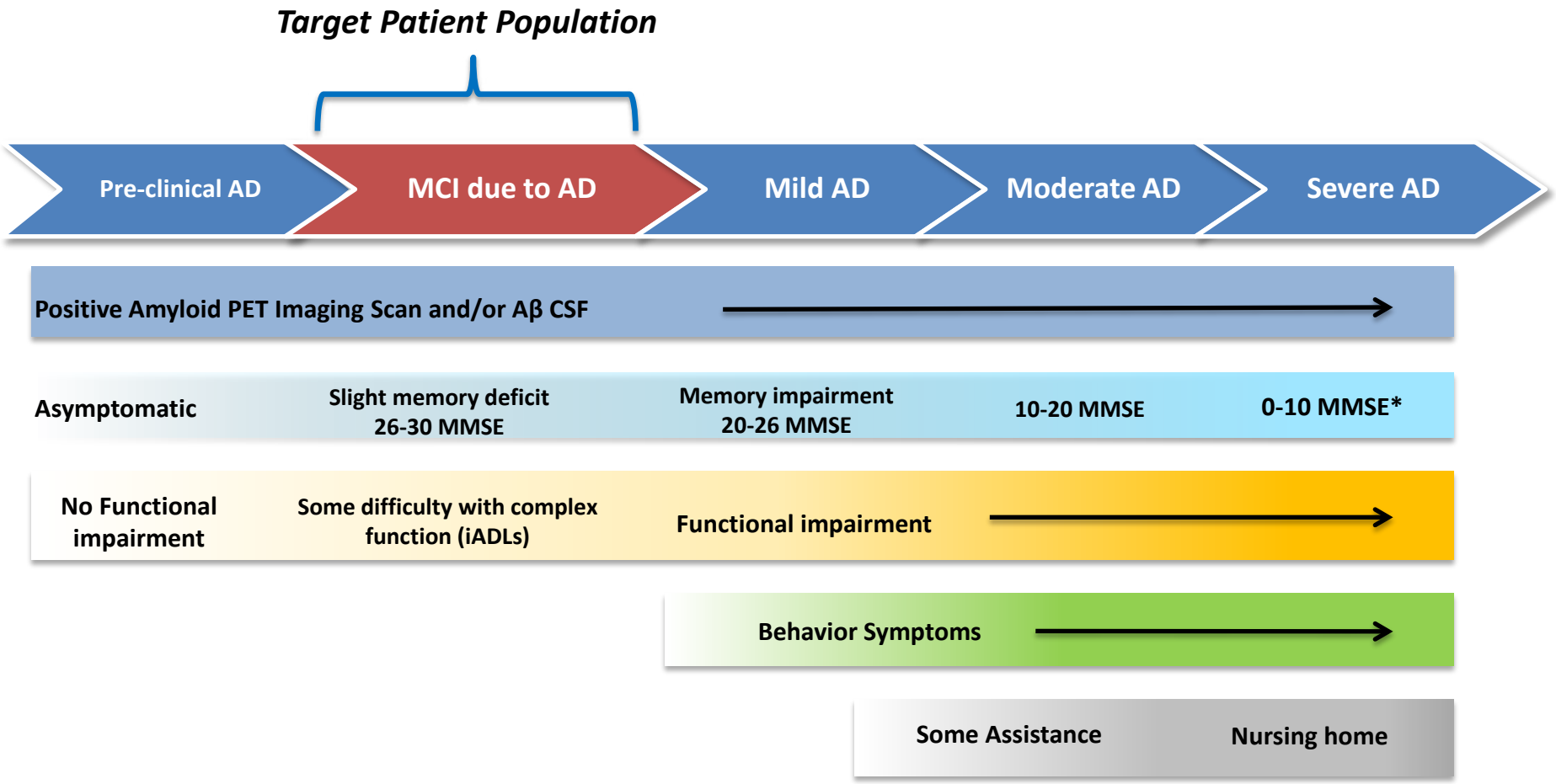
53

Assessing Performance Outcome Measures in Mild Cognitive Impairment due to Alzheimer's Disease: A C-Path Case Study

J. Scott Andrews, PharmD

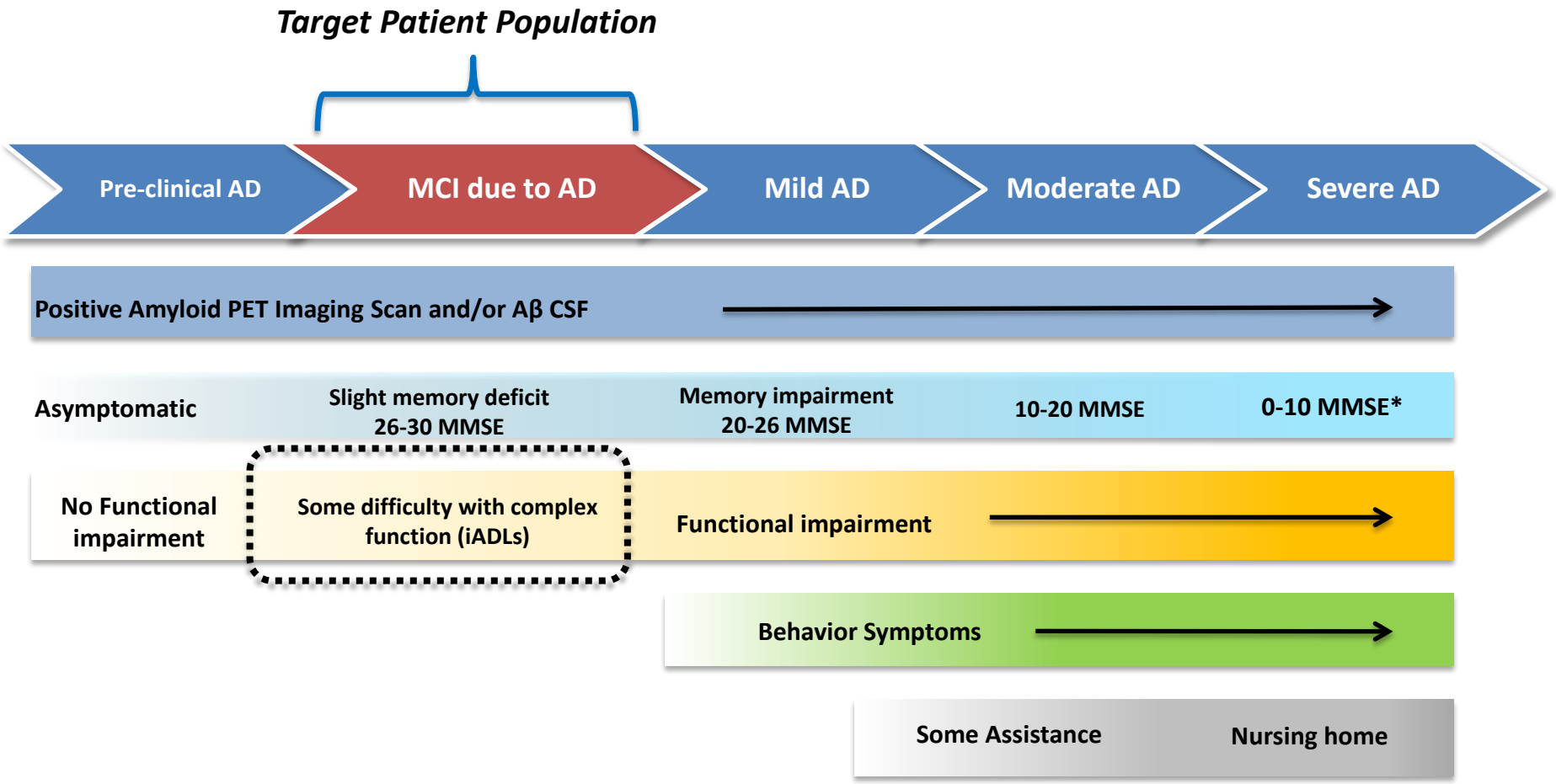
Research Scientist, Eli Lilly and Company

Measuring Function Across the Continuum of Alzheimer's Disease



*MMSE cutoffs representative

Measuring Function Across the Continuum of Alzheimer's Disease



*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 ?**

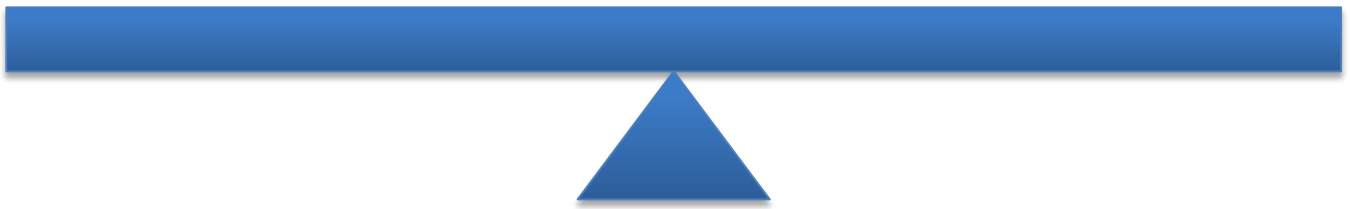
PerfO

Opportunities

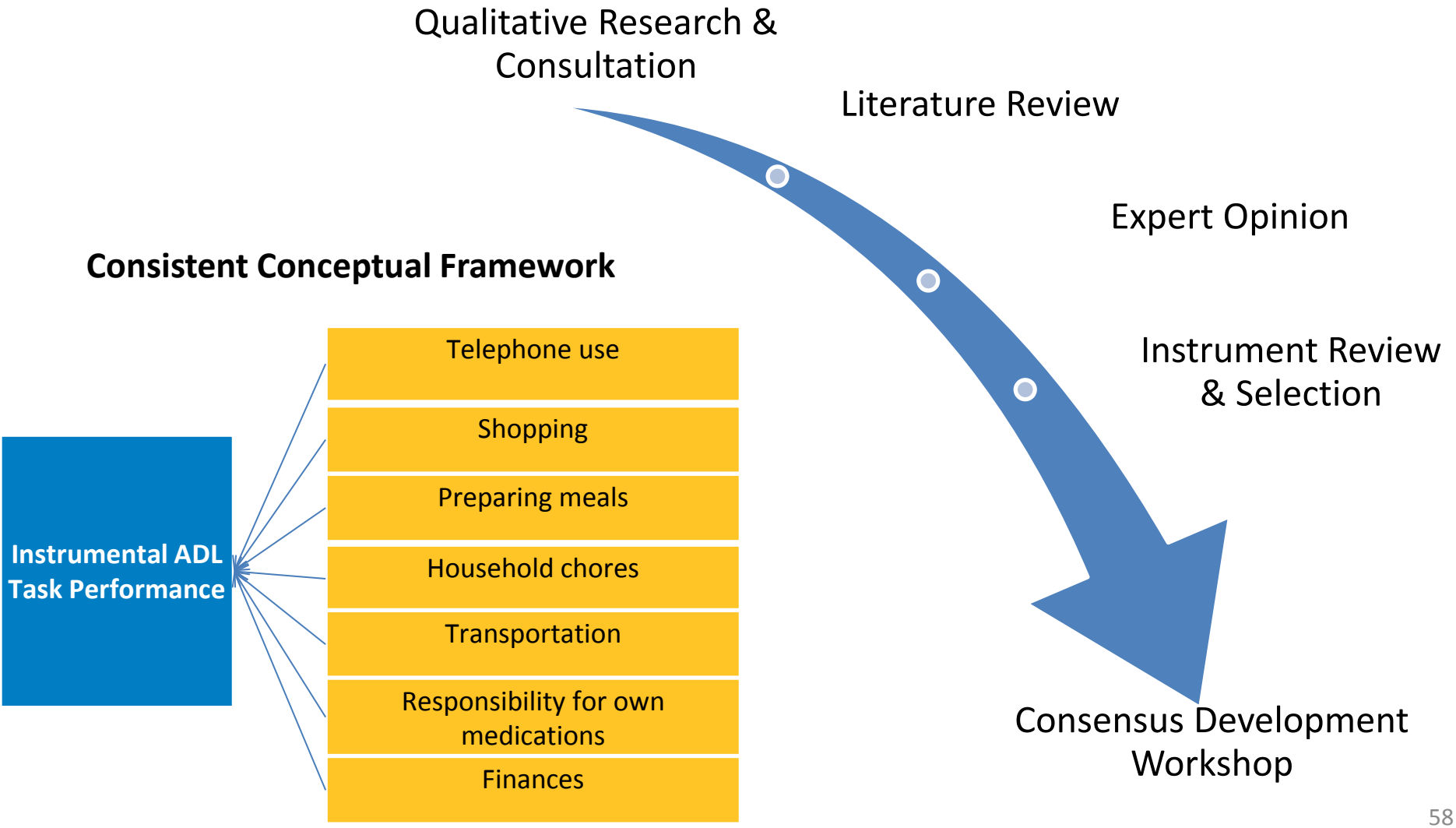
- Measurement properties
- Patient-focus
- Direct evidence?

Challenges

- Operational feasibility
- Cross-cultural applicability
- Content validity?
- Interpretation?

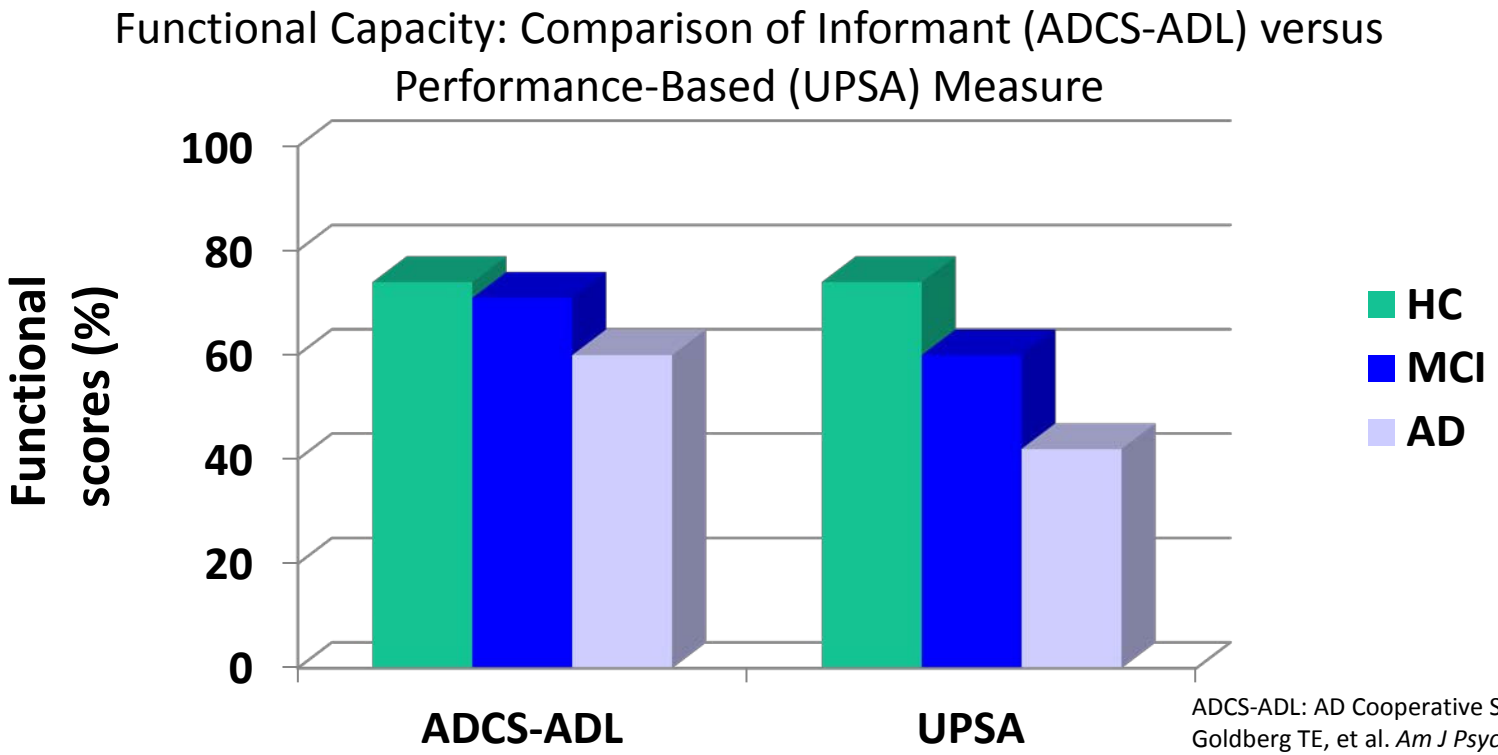


Pathway for Measurement Selection



PerfO Instrument Review

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



ADCS-ADL: AD Cooperative Study ADL Scale

Goldberg TE, et al. *Am J Psychiatry* 2010;167:845-853

Consensus Development Feedback

Attendees:

- FDA Division of Neurology Products
- FDA Clinical Outcome Assessment
- FDA Office of Biostatistics
- Expert Consultants
- C-Path
- Adelphi Values
- Industry Members

Outcome: explore qualification of UPSA

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

- FDA DNP

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

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.
- *J. Scott Andrews, PharmD* – Research Scientist, Eli Lilly and Company and Co-chair PRO Consortium's Cognition WG

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?

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

[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

Review Division	Disease/Condition	Indication and/or Claim(s) Description	Outcome of Interest	COA (COA Type)	COA Context of Use	COA Qualification Information
TRANSPLANT AND OPTHAMOLOGY PRODUCTS	Neovascular (wet) age-related macular degeneration	Treatment of age-related macular degeneration	Best corrected visual acuity	Visual acuity (performance outcome)	Adult patients with age-related macular degeneration	Not applicable
CARDIOVASCULAR AND RENAL PRODUCTS	Chronic thromboembolic plumonary hypertension (CTEPH)	Treatment of persistent/recurrent CTEPH after surgical treatment or inoperable CTEPH to improve exercise capacity and WHO functional class	Exercise capacity	6-Minute Walking Distance (performance outcome)	Adult patients with CTEPH	Not applicable
CARDIOVASCULAR AND RENAL PRODUCTS	Pulmonary arterial hypertension	Treatment of pulmonary arterial hypertension	Exercise capacity Incidence of death or clinical deterioration	6-minute Walking Distance (performance outcome) Survival and/or clinician-reported outcome	Adult patients with pulmonary arterial hypertension	Not applicable
NEUROLOGY PRODUCTS	Alzheimer's disease: Mild cognitive impairment due to Alzheimer's disease (MCI due to AD)	To be determined	Day-to-day functioning (instrumental activities of daily living)	Currently unnamed (performance outcome tool to assess instrumental activities of daily living (IADLs))	Adults (≥45 years with mild cognitive impairment due to Alzheimer's disease (MCI due to AD)	Submitter: Critical Path Institute: PRO Consortium's Cognition Working Group

Some PerfO Assessments included in the FDA's pilot COA Compendium (cont.)

Review Division	Disease/Condition	Indication and/or Claim(s) Description	Outcome of Interest	COA (COA Type)	COA Context of Use	COA Qualification Information
NEUROLOGY PRODUCTS	Multiple Sclerosis (MS)	To be determined	"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	New Clinical Outcome Assessment Instrument for Use in Clinical Trials of Medical Products to Treat Multiple Sclerosis (MS) (performance outcome)	Adults living with relapsing-remitting or progressive forms of MS	Submitter: Critical Path Institute Multiple Sclerosis Outcome Assessments Consortium (MSOCAC)
GASTROENTEROLOGY AND INBORN ERRORS PRODUCTS	Mucopolysaccharidosis I (MPS I) (Hurler and Hurler-Scheie forms of MPS I)	Improvement in walking capacity	Walking capacity	6-Minute Walk Test (performance outcome)	Pediatric and/or adult patients with MPS I	
METABOLISM AND ENDOCRINOLOGY PRODUCTS	Muscle wasting disorder (lower extremity functional decline in patients with hip fracture)	To be determined	Lower-extremity functional decline	Usual Gait Speed (UGS) and the Short Physical Performance Battery Test (SPPB) (performance outcome)	Persons age 65 years and older who have diminished muscle mass and strength and decreased function that is a result of a hip fracture	Submitter: Aging in Motion Coalition of the Alliance for Aging Research

VRFCAT Future Updates

Enhanced study and subject set-up

Graphical Upgrade – Unreal Game Engine 4.0

Protections for data from deletion/editing, enhanced security

Audit Trail Functionality

Required user credentials and Fingerprint Scanning for Subject Verification

'Clinical Trial Mode' – Locking settings, etc.

Provide sites read-only access to data on server

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

		Healthy Controls			Schizophrenia Patients				
		N	M	SD	N	M	SD		
								t	p
Age		165	42.6	13.94	158	43.6	11.85	-0.72	.47
Years of Education		165	14.7	2.41	157	12.8	1.99	7.77	<.001
Mother's Education		155	12.9	2.98	142	12.5	3.33	1.18	.24
		% (N)			% (N)			X^2	p
Male		53 (88)			56 (88)			0.18	.67
Unemployed		33 (54)			85 (135)			92.40	<.001
Comfortable with Computer		97 (160)			89 (140)			8.53	.004
Hispanic		18 (29)			19 (30)			0.11	.74
English Primary Language		95 (157)			96 (151)			0.03	.86
Race									
	Caucasian	56 (92)			47 (75)			3.33	.19
	African American	38 (63)			48 (76)				
	Other	6 (10)			4 (7)				

Results: Visit 1 data for SZ group

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

Clinician SCoRS Total, Mean (SD)	38.2 (9.88)
PANSS Total, Mean (SD)	71.6 (21.93)
Clinician SLOF Total, Mean (SD)	120.8 (14.42)

- Schizophrenia Cognition Rating Scale (SCoRS)
- Positive and Negative Syndrome Scale (PANSS)
- Specific Level Of Functinoing (SLOF) Total Scores

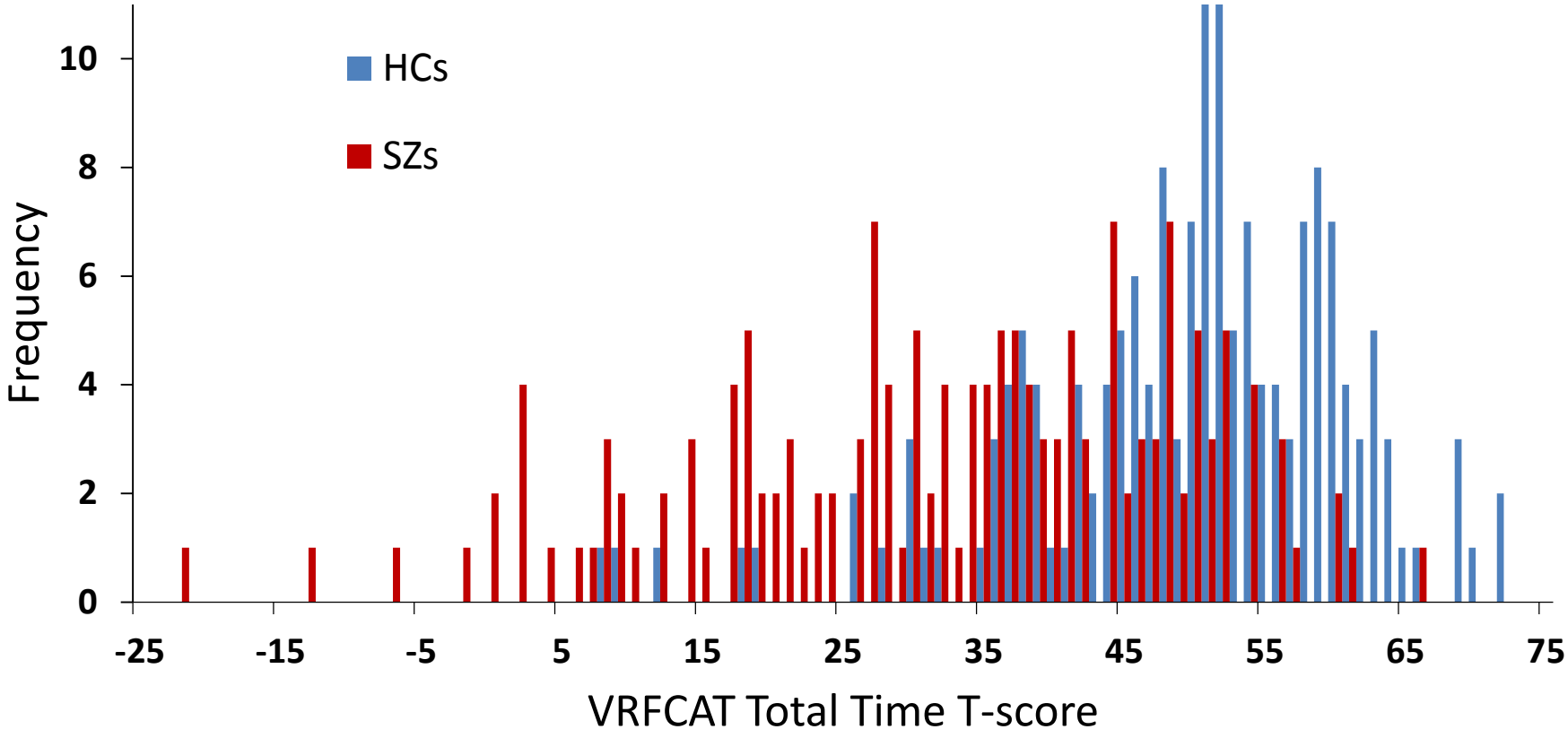
Results: Ability to Discriminate Between SZs and HCs

- 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

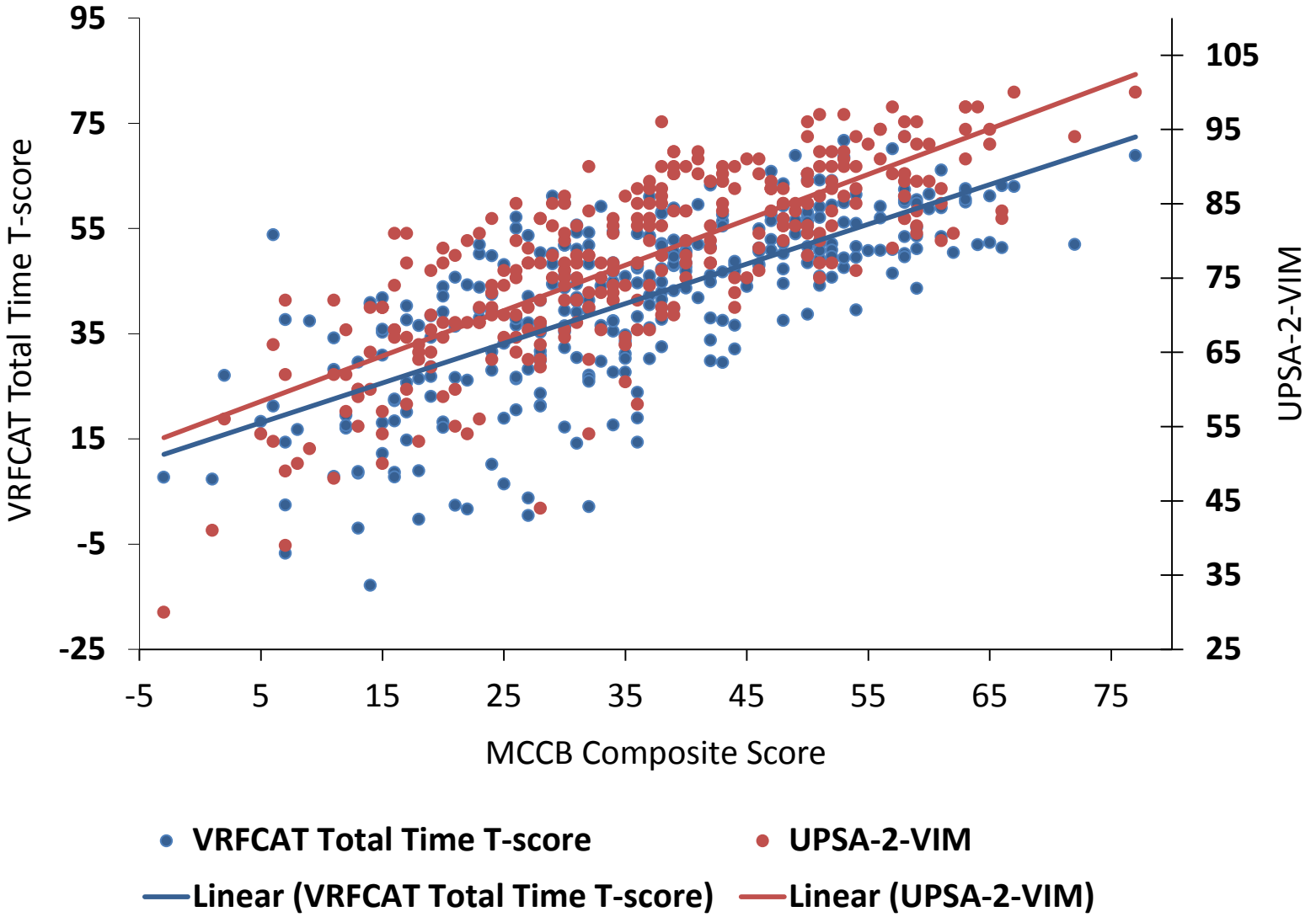
	HC (N = 165)	SZ (N = 158)	Cohen's d
MCCB Composite Score, Mean (SD)*	44.0 (13.19)	28.1 (12.91)	1.22
VRFCAT Total Time T-score, Mean (SD)*	49.7 (11.51)	32.5 (16.60)	1.21
VRFCAT Total Errors T-score, Mean (SD)*	49.4 (11.62)	37.6 (22.37)	0.67
VRFCAT Progression T-score, Mean (SD)*	49.7 (10.16)	40.5 (13.62)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

* Indicates significant differences between HC and SZ at the 0.05 significance level.

Results: VRFCAT Total Time T-score Discrimination



Results: Comparison of Performance on Different Measures at Visit 1



Results: Evaluating Age Differences in Healthy Population

Functional capacity and cognitive performance by age

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

- 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)

Study Conclusions

- 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 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

Assessment	Visit 1 Mean (SD)		Visit 2 Mean (SD)		Difference Mean (SD)		Cohen's d[1]		Intraclass Correlation Coefficient (ICC)	
	YA	OA	YA	OA	YA	OA	YA	OA	YA	OA
VRFCAT Total Time (minutes)	11.8 (2.10)	14.6 (2.52)	11.5 (2.25)	14.3 (3.45)	0.3 (1.38)	0.3 (2.56)	0.1	0.1	0.80	0.64
VRFCAT Total Errors	1.1 (1.46)	2.8 (3.04)	0.9 (1.28)	2.8 (4.65)	0.2 (1.43)	0.0 (4.19)	0.1	0.0	0.46	0.44
VRFCAT Total Progressions	0.0 (0.22)	0.5 (0.72)	0.0 (0.22)	0.4 (0.93)	0.0 (0.22)	0.1 (0.82)	0.0	0.1	0.48	0.52
Total Bus Schedule Checks	3.4 (1.91)	3.5 (1.67)	3.1 (1.70)	3.8 (2.39)	0.3 (2.00)	-0.3 (2.49)	0.2	-0.1	0.39	0.28
Total Recipe Checks	12.5 (4.91)	11.9 (5.10)	12.2 (5.46)	11.9 (5.23)	0.3 (4.50)	0.0 (3.55)	0.1	0.0	0.63	0.77
UPSA-2-VIM	84.8 (8.45)	83.4 (8.98)	87.6 (8.21)	86.2 (9.56)	2.7 (5.76)**	2.7 (6.58)*	0.3	0.3	0.72	0.72

- 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).