

***Welcome to
Ninth Annual
Patient-Reported Outcome
Consortium Workshop***

April 25 – 26, 2018 ■ Silver Spring, MD

Network: Cypress-Guest ; Passcode: Cyp24689



Why reinvent the wheel?

***Ninth Annual
Patient-Reported Outcome Consortium Workshop***

April 25 – 26, 2018 ■ 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 or the Critical Path Institute.
- 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 Outline



- Industry Perspective: Measurement approaches and evidentiary expectations for using or modifying existing instruments
- FDA Perspective: Approaches to modification and evidentiary considerations
- Use “as is”: *PROMIS[®]/FACIT Fatigue Scale* as examples
- Modification: *SMDDS* as example
- Panel Discussion
- Q & A

Session Participants



Moderator

- *Maria Mattera, MPH* – Assistant Director, Patient-Reported Outcome Consortium, C-Path

Presenters

- *Elizabeth (Nicki) Bush, MHS* – Director, Patient-Focused Outcomes Center of Expertise, Eli Lilly and Company, and Industry Co-Director, PRO Consortium
- *Elektra Papadopoulos, MD, MPH* – Associate Director, COA Staff, OND, CDER, FDA
- *Dave Cella, PhD* – Professor and Chair, Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University
- *Sonya Eremenco, MA* – Associate Director, Patient-Reported Outcome Consortium, C-Path

Panelists

- *Billy Dunn, MD* – Director, Division of Neurology Products, OND, CDER, FDA

Why Reinvent the Wheel? One Industry Perspective

Elizabeth (Nicki) Bush, MHS

Director, Patient-Focused Outcomes Center of Expertise

Eli Lilly and Company

Measurement Strategy Considerations



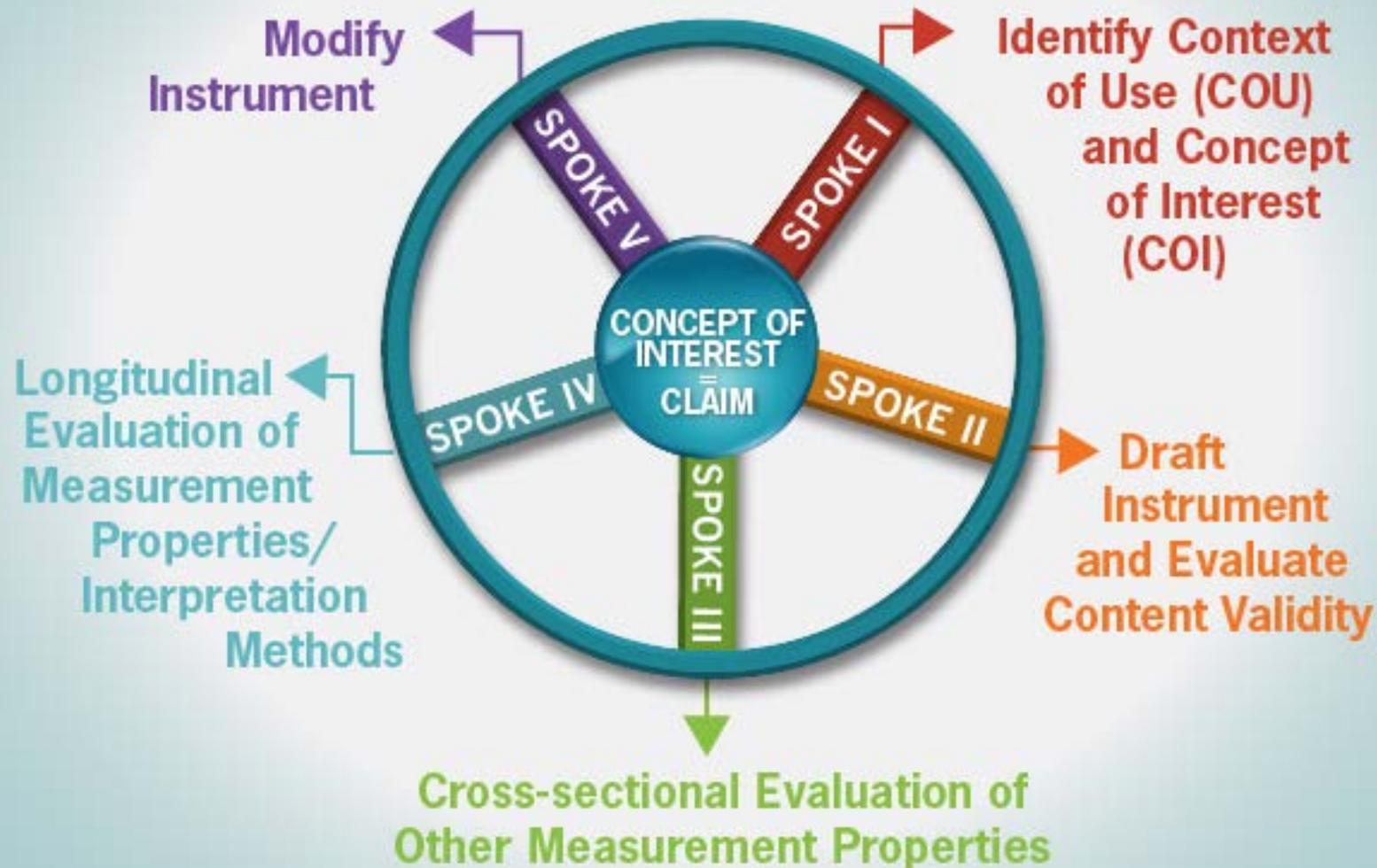
- What is/are the concept(s) of interest?
 - Literature, patients, clinicians
- What is the context of use?
 - Planned study design
- How will the data be used?
 - Communications? Publications? Label? Benefit/Risk characterization?
- Landscape
 - Regulatory precedent
 - Other labels
- All in context of drug development

Does a Fit-for-purpose Tool Exist?

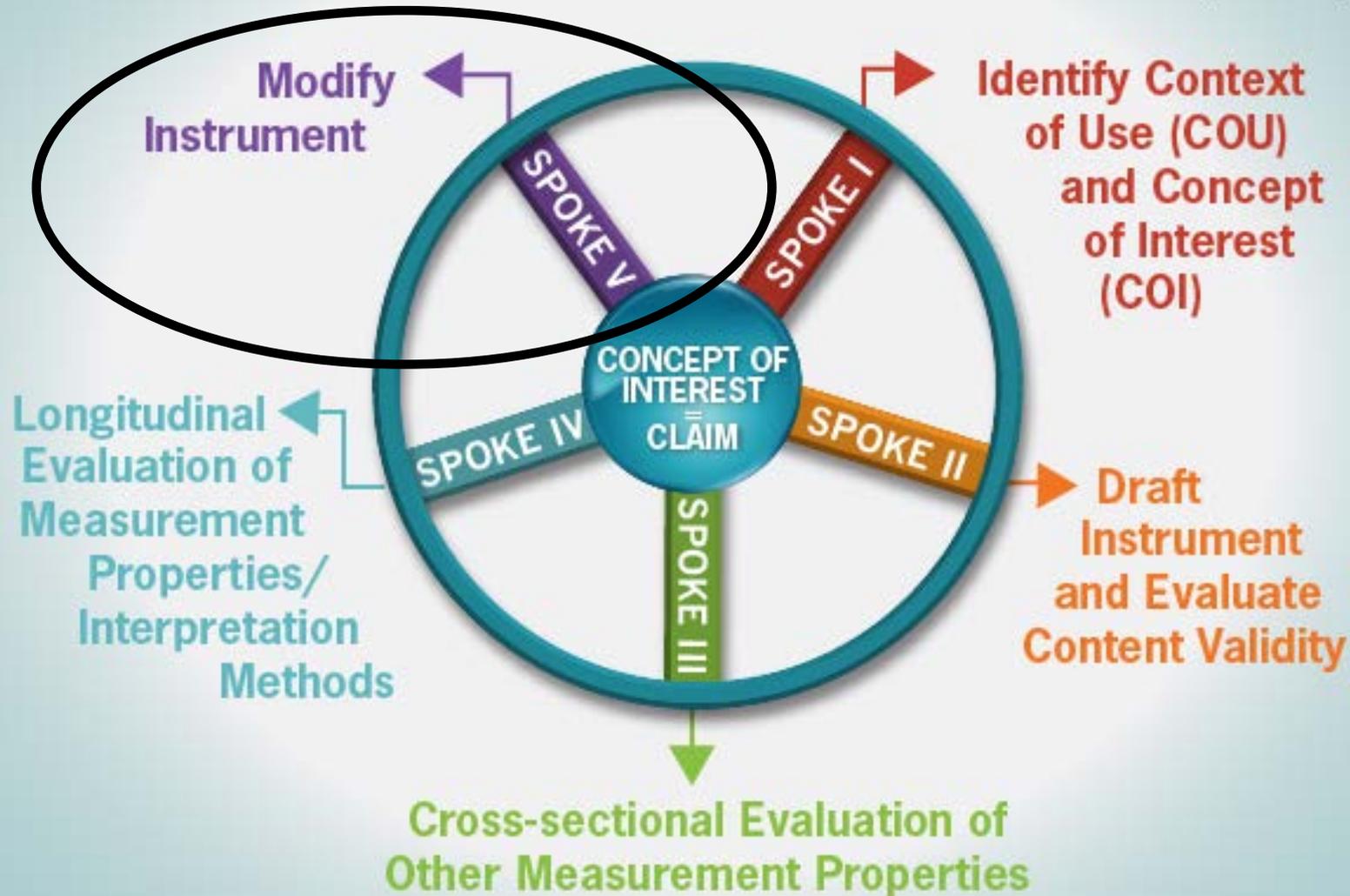


- Identify potential tools
- Assess: conceptual overlap; development; content validity and other measurement properties
- Identify/Address concerns
 - Documentation of development, content validity, other measurement properties
 - If modifications are needed
 - Will the instrument developer be amenable?
 - How much of the original development and validation evidence can be used to support the new version?
 - How much change is too much?
 - Regulatory

CLINICAL OUTCOME ASSESSMENTS (COAs)



CLINICAL OUTCOME ASSESSMENTS (COAs)



Should We Develop a New Tool?



- De novo development often perceived to be “safer” in labeling discussions
 - Control over all aspects of development (and the documentation)
- There *are* some risks to de novo development
 - Lack of familiarity/trust (researchers, regulators, prescribers)
 - Interpretability
 - Measurement properties not well-established
 - Newer is not always better
- Path to label seems clearer—or at least more straightforward

Let's Discuss



- How does the FDA *really* feel about existing tools?
- What evidence for existing tools is expected? What makes sense?

Let's Discuss



- How does the FDA *really* feel about existing tools?
- What evidence for existing tools is expected? What makes sense?

Today's discussion will *not* touch on modifications made due to translations or migration to another mode of administration

PRO Consortium 9th Annual Meeting: Why Reinvent the Wheel?

*Use or modification of existing
instruments*

Elektra J Papadopoulos, MD, MPH
Associate Director for Clinical Outcome
Assessments Staff; Office of New Drugs;
FDA Center for Drug Evaluation and Research

April 26, 2018

Disclaimer

The views and opinions expressed in this presentation are those of the individual presenter and should not be attributed to the U.S. Food and Drug Administration (FDA).

Fit-for-purpose*

- For medical product development tools, fit-for-purpose is a conclusion that the level of validation associated with a tool is sufficient to support its context of use

*BEST (Biomarkers, EndpointS, and other Tools) Resource
<https://www.ncbi.nlm.nih.gov/books/NBK338448/>

What makes a COA “fit-for-purpose” for medical product development?

- Appropriate for its intended use e.g.,
 - Study design
 - Patient population
- Validly and reliably measure a concept that is
 - Clinically relevant
 - Important to patients
- Can be communicated in labeling in a way that is accurate, interpretable, and not misleading (i.e., well-defined)*

* If the COA is appropriately applied in medical product development

Importance of fit-for-purpose COAs: Why do we care?

- Use of an inadequately developed or tested instrument introduces risk whether the instrument is used “off-the-shelf,” modified, or developed *de novo*
- Lack of a thoughtful approach to measurement may lead to:
 - (1) Content validity problems: Misleading content such that the tool does not accurately assess target concept in the target population (may compromise to ability to accurately describe clinical benefit in labeling)
 - (2) Poor ability to detect change (may compromise ability to detect a treatment effect when one exists)

The challenge

- With the many thousands of diseases and potential drugs in need of development it's impossible to have an instrument fully developed and evaluated specifically for each disease and context of use, so we often look to utilize existing COAs
- The decision is whether:
 - An existing tool is fit for purpose for the new context of use;
 - An existing tool could be modified for the new context of use; or
 - A new tool should be developed “from scratch”

Common challenges with use of existing PRO instruments

- Development history may not be available
- Development lacking patient input
- Developed for purposes other than clinical trials/medical product development

Modification or use of existing COAs: Regulatory considerations

Content validity evaluation for PROs:



Use of existing instruments for a new context*

- 1. Conceptual match:** Does the concept measured by the instrument match the concept targeted for a specific claim in the population? If not what are the differences? Could a relevant portion of an existing instrument be identified and used?
- 2. Input from target patient population:** Does the patient population in which the PRO measure was developed compare well with that of the target patient population for the clinical trial? If not, what are the differences? What additional data might be needed to supplement what is known?
- 3. Item content:** To what degree do the items adequately cover the concept? Are irrelevant items included? Are important items for that concept and population omitted?
- 4. Modifications:** What modifications needed based on consideration from 1-3 above?

What does FDA want to see?

- It's a misconception that FDA only considers a PRO “fit-for-purpose” if every box in the PRO guidance has been checked
- Instead, FDA encourages a thoughtful and strategic approach to instrument development using what is already known from literature, expert input, and patient input and obtaining additional information as needed to ensure that:
 - (1) Concepts important to the target patient population and drug claim are covered
 - (2) Patients understand the items as intended
 - (3) Score is able to show change in the clinical trial context of use

Potential Applications of an Existing Instrument

- Context of original development and validation
- New patient population
- Same population, but new clinical trial design

Uses of existing PRO measures

- No modifications to the instrument, use in original context of use
- No modification to the instrument, but application in a new context of use
- Modification to the instrument for a new context of use
 - Instructions/training materials
 - Recall period
 - Item content (e.g., dropping, adding or modifying wording of items)
 - Change of item stems
 - Change of response options
 - Item order
 - Scoring
 - Formatting

*Mode of administration (paper, electronic)—not addressed in today's session

Evidentiary considerations

- Based on the magnitude of the modification, evidence of fitness-for-purpose can be developed through:
 - Concept confirmation
 - Cognitive interviewing
 - Full psychometric testing as an additional step (if substantial modifications have been made)

Example:

Use of an existing measure across the age continuum

- Adults → children
 - Conceptual match: Is the disease the same in adults as it is in children or is the experience different such that the concepts relevant to adults do not apply in children? If the latter, de novo instrument development would be needed
 - If the concepts are similar across the ages, it may be possible to adapt an existing PRO measure so that it is more age appropriate for use in children
 - Depending on the age and cognitive ability, children may require more simple concrete concepts, fewer response options, pictorial scales, shorter recall period
 - If the target population of children cannot self-report, an observer-reported outcome would be needed and de novo instrument development would be needed

Summary

- FDA evaluates a COA to ensure it's fit-for-purpose in medical product development
- Careful thought and planning is needed to ensure that existing tools are appropriately applied in drug development to avoid compromising measurement properties including content validity
- FDA is open to working with sponsors to aid in decision making with regard to use or modification of existing instruments for a new context of use



U.S. FOOD & DRUG
ADMINISTRATION

Use “as is”

PROMIS/FACIT Examples of Avoiding Wheel Re-invention

David Cella, PhD

Northwestern University, Chicago, IL, USA

Overview

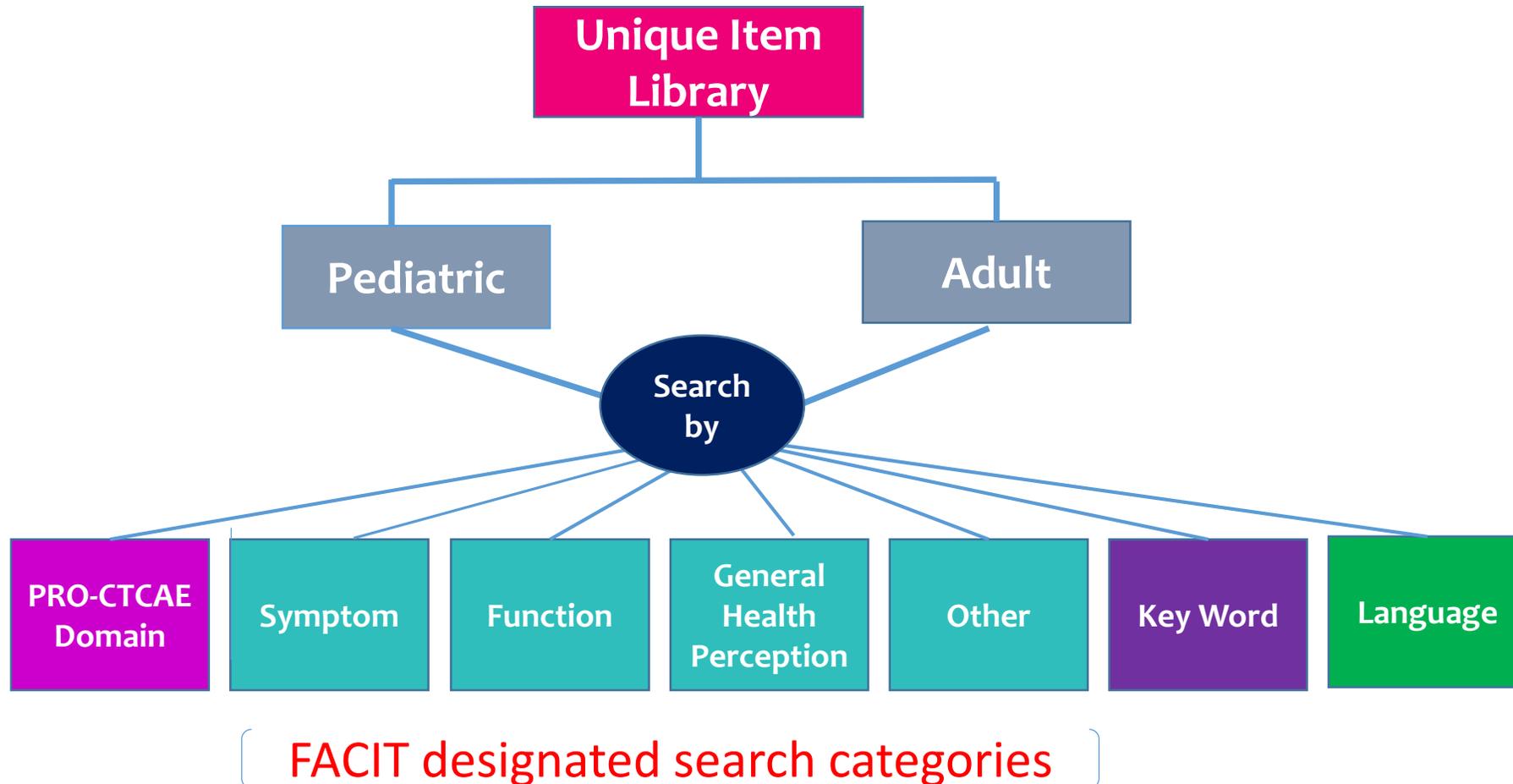
- Defining “as is” --- not *exactly* as is
- PROMIS and FACIT-Fatigue: How they  a domain
- Physical Function: From *unfit*  to *fit*  for purpose
- Some related issues
 - Content validity
 - Recall period
 - An illustration in Multiple Sclerosis

Defining “as is” --- not *exactly* as is



- Gen 1: Generic Developed 1970 to 1990
 - Gen 2: Disease-specific Developed 1990 to 2010
 - Gen 3: Modular/Custom Emerging 2010 →
-
- Gen 1 to 2 tools tend to be truly as-is (and not bad when re-purposed)
 - Gen 3 tools designed to accommodate study-specific needs
 - “As-is” redefined
 - PRO-CTCAE, FACIT Searchable Item Library (SIL), EORTC Library – pick and choose
 - PROMIS and other **calibrated item banks** (a whole new way)

Gen 3 example: FACIT-Searchable Item Library



Item Banks

- Collection of items (questions)
- All measure the same thing
- Each item stands alone as a calibrated component of the bank
- This opens up special opportunities to build:
 - “Well-defined” measures of common symptoms and functional abilities
 - Fit for regulatory purposes across diseases and treatment contexts



PROMIS T-Score

- Mean = 50
 - Standard Deviation = 10
- Referenced to the US General Population
- High scores = more of domain
- MID = 2 to 6 points (physical function)^{1,2}



¹Yost 2011, ²Hays 2015



HealthMeasures ₃₅
TRANSFORMING HOW HEALTH IS MEASURED

Domains and Item Banks

A **domain** is the specific feeling, function or perception you want to measure.

Cuts across different diseases

Examples

- Fatigue
 - Pain
 - Anxiety
 - Depression
- Physical function
 - Cognitive function
 - Sleep/Wake function
- Global health perceptions
 - Satisfaction with social participation

PROMIS and FACIT-Fatigue: How they HARMONIZE a domain



- PROMIS has a 95-item fatigue bank
- 13 of them are FACIT-Fatigue items
- Unidimensional with wide range of content
- Therefore, FACIT-F is also a 13-item PROMIS short form

PROsetta Stone® Crosswalk

FACIT-F to PROMIS Fatigue

FACIT-F	PROMIS T	SE	FACIT-F	PROMIS T	SE
52	30.3	4.8	12	68.9	2.0
51	35.0	3.5	11	69.6	2.0
50	38.0	3.0	10	70.4	2.0
49	40.3	2.8	9	71.2	2.1
48	42.1	2.6	8	72.0	2.2
47	43.7	2.5	7	72.9	2.3
46	45.0	2.3	6	73.9	2.4
45	46.3	2.2	5	75.0	2.5
44	47.3	2.1	4	76.2	2.7
43	48.3	2.0	3	77.5	2.9
42	49.3	2.0	2	79.1	3.1
41	50.1	1.9	1	81.2	3.3
40	51.0	1.9	0	83.5	3.4
39	51.7	1.9			
38	52.5	1.9			
37	53.2	1.9			
36	53.9	1.8			
35	54.6	1.8			
34	55.3	1.8			
33	55.9	1.8			
32	56.6	1.8			
31	57.2	1.8			
30	57.8	1.8			
29	58.4	1.8			
28	59.0	1.8			
27	59.6	1.8			
26	60.2	1.8			
25	60.8	1.8			
24	61.4	1.8			
23	62.0	1.8			
22	62.6	1.8			
21	63.2	1.8			
20	63.8	1.8			
19	64.4	1.8			
18	65.0	1.8			
17	65.6	1.8			
16	66.2	1.9			
15	66.9	1.9			
14	67.5	1.9			
13	68.2	1.9			

www.prosettastone.org

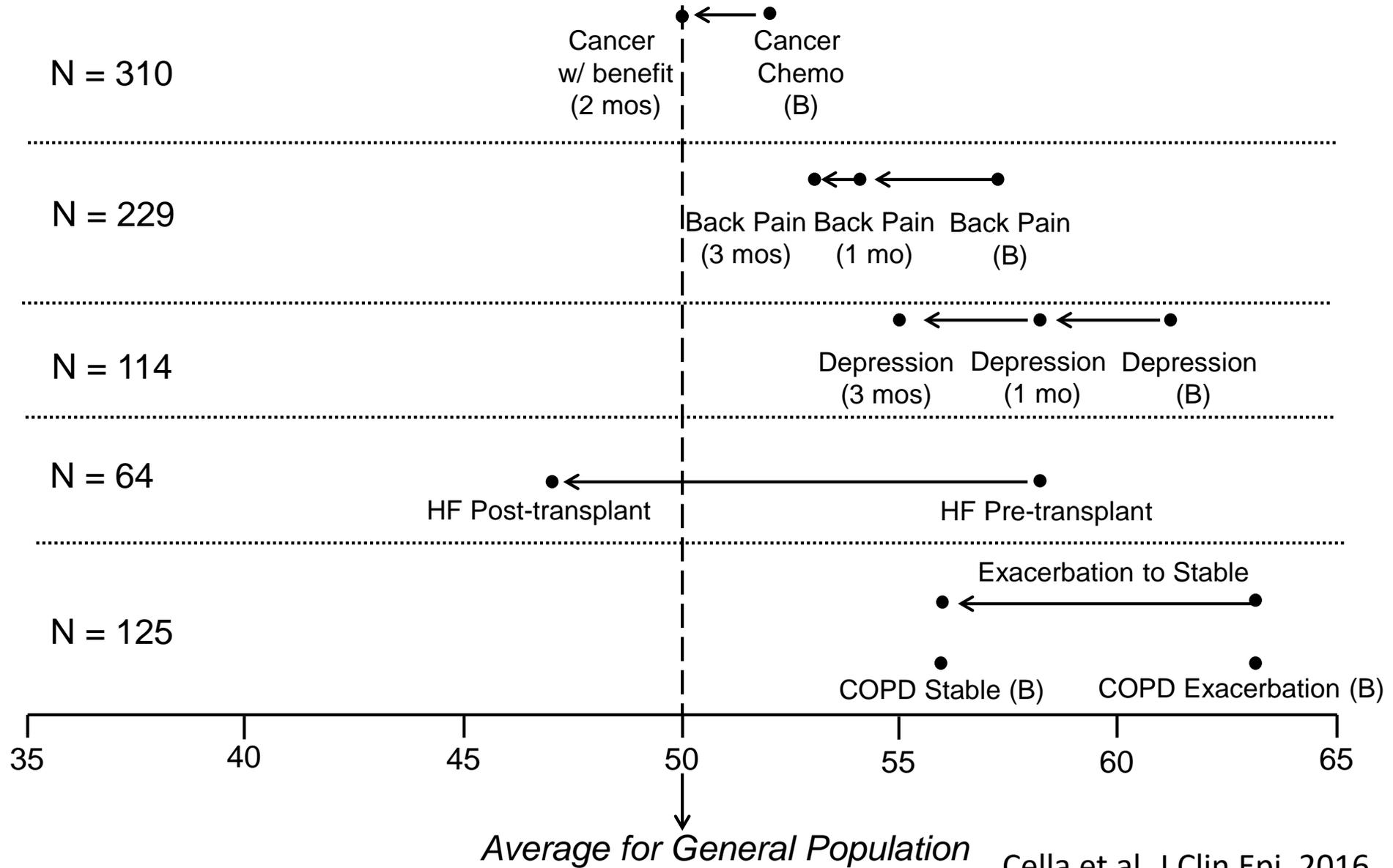
Conditions where FACIT-F has been used/validated

cancer – adult, on and off chemotherapy
cancer- pediatric, on and off chemotherapy
rheumatoid arthritis
osteoarthritis
psoriatic arthritis
ankylosing spondylitis
multiple sclerosis
psoriasis
systemic sclerosis/scleroderma
sarcoidosis
paroxysmal nocturnal hemoglobinuria
chronic obstructive pulmonary disease (COPD)
carnitine deficiency

myelodysplastic syndrome
HIV/AIDS
systemic lupus erythematosus (SLE)
hidradenitis suppurativa
Crohn's disease
ulcerative colitis
Sjögren's syndrome
Gaucher disease
chronic immune thrombocytopenia
chronic kidney disease
Parkinson's disease
stroke

...and a US general population reference sample

PROMIS Fatigue Across Five Clinical Conditions



PROMIS Physical Functioning on the FDA COA Compendium: Undergoing COA Qualification in Oncology

PILOT
CLINICAL OUTCOME ASSESSMENT COMPENDIUM
(COA Compendium)
Version 1
January 12, 2016



Information Based on Drug Labeling Approved From 2003 to 2014: December 31, 2014; and
 CDER's DDT COA Qualification Program: December 31, 2015

ONCOLOGY PRODUCTS 1

Disease/Condition	Indication and/or Claim(s) Description ^{38 39}	Outcome of Interest	COA (COA Type)	COA Context of Use	COA Qualification Information
Prostate cancer (metastatic castration-resistant) Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics	Improvement in pain or delay in time to pain progression for patients with metastatic castration-resistant prostate cancer	Pain intensity	Brief Pain Inventory Item #3 – Short Form (patient-reported outcome)	Use in adult patients with metastatic castration-resistant prostate cancer along with other key oncology measures (e.g., overall survival)	Not applicable
Physical functioning in oncology	To be determined	Physical functioning	Patient Reported Outcome Measurement System (PROMIS) – Physical Function item bank (patient-reported outcome)	For use in clinical trials in oncology	Submitter: PROMIS Network Center Visit “Clinical Outcome Assessment Qualification Program Submissions” Web site for additional information

From *unfit* to *fit* for purpose



FACT-Physical Well-being: Getting to “well-defined”

~~1. I have a lack of energy~~

~~2. I have nausea~~

3. Because of my physical condition, I have trouble meeting the needs of my family

~~4. I have pain~~

~~5. I am bothered by side effects of treatment~~

~~6. I feel ill~~

7. I am forced to spend time in bed

PROMIS Items to Add to FACT-G PWB-2

- PFA1 - Does your health now limit you in doing vigorous activities, such as running, lifting heavy objects, or participating in strenuous sports?
- PFA11 - Are you able to do chores such as vacuuming or yard work?
- PFA53 - Are you able to run errands and shop?

Combining PROMIS PF and FACT-PWB

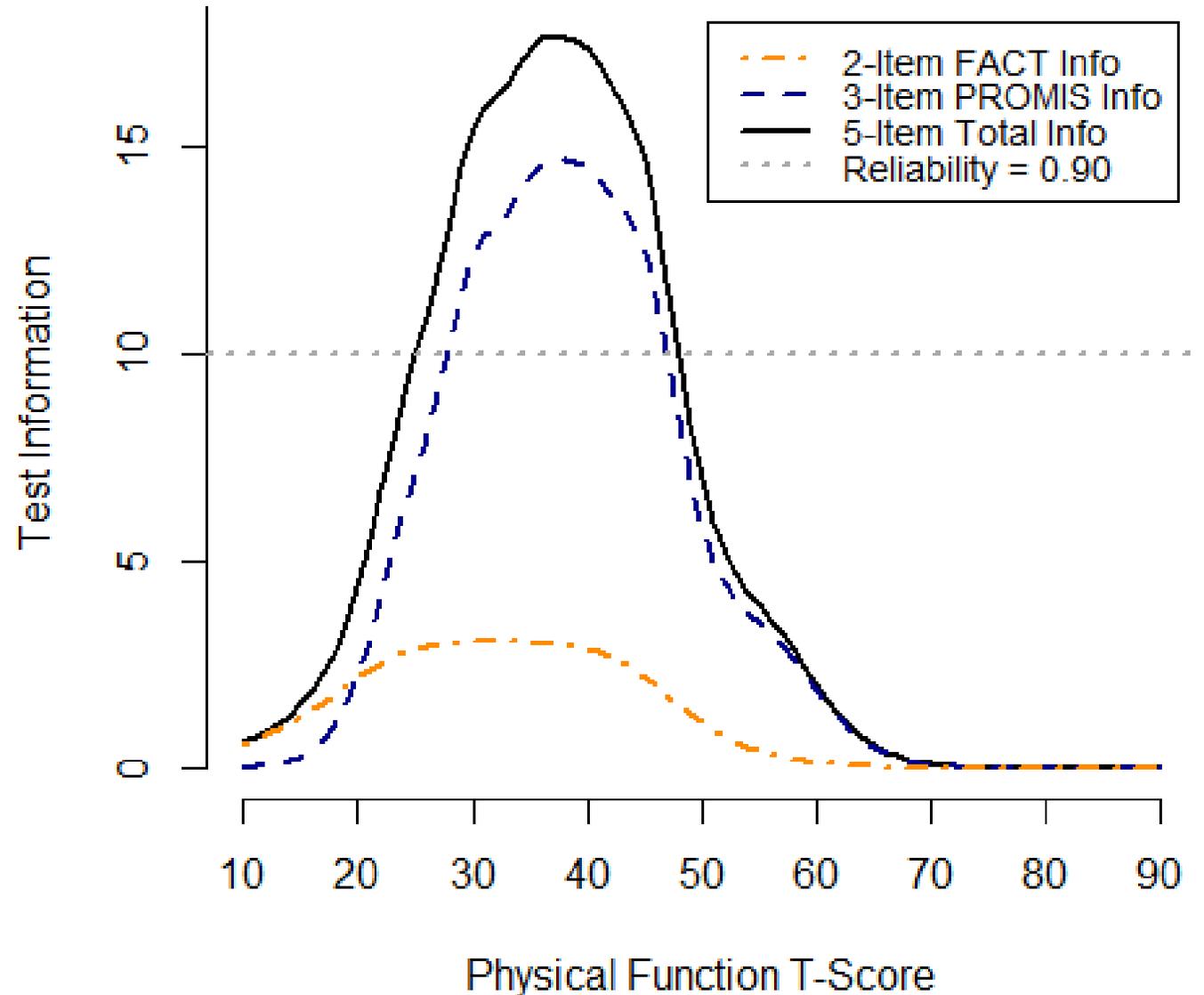
From unfit



To fit



FACT-PROMIS-PF5 Test Information



Some Related Issues



- Content validity
- Recall period
- An illustration in Multiple Sclerosis

Content Validity



A well-defined domain, with flexible assessment options, frees us up to break from a disease-oriented approach to measuring common symptoms and function

Does Recall Period Even Matter? A Study of Cancer Patients and the General Population (GP)

N=1,000 Cancer; 1,400 GP, Random assignment

		Recall period		
		24 hours	7 days	None
Recall placement	With every item	24/yes n=200	7/yes n=200	
	At beginning only	24/no n=200	7/no n=200	
	None			0/no n=200 Cancer n=600 GP

Differences in PROMIS-Physical Function T-Score Means compared to the “0/no” group (PROMIS standard)

	Cancer 31-item	Cancer 10-item	Gen Pop 31-item	Gen Pop 10-item
7/no	1.44	1.16	-0.73	-0.86
7/yes	1.02	0.62	0.49	0.45
24/no	1.63	1.49	0.23	0.02
24/yes	2.30	2.11	2.15	2.01
P-value	0.052	0.056	0.031	0.020

Estimated mean differences from models adjusting for age, gender, education, race/ethnicity, English fluency, and presence of comorbidities (and diagnosis time, cancer and treatment type for cancer group)

P-value is for overall F test for effect of randomized group

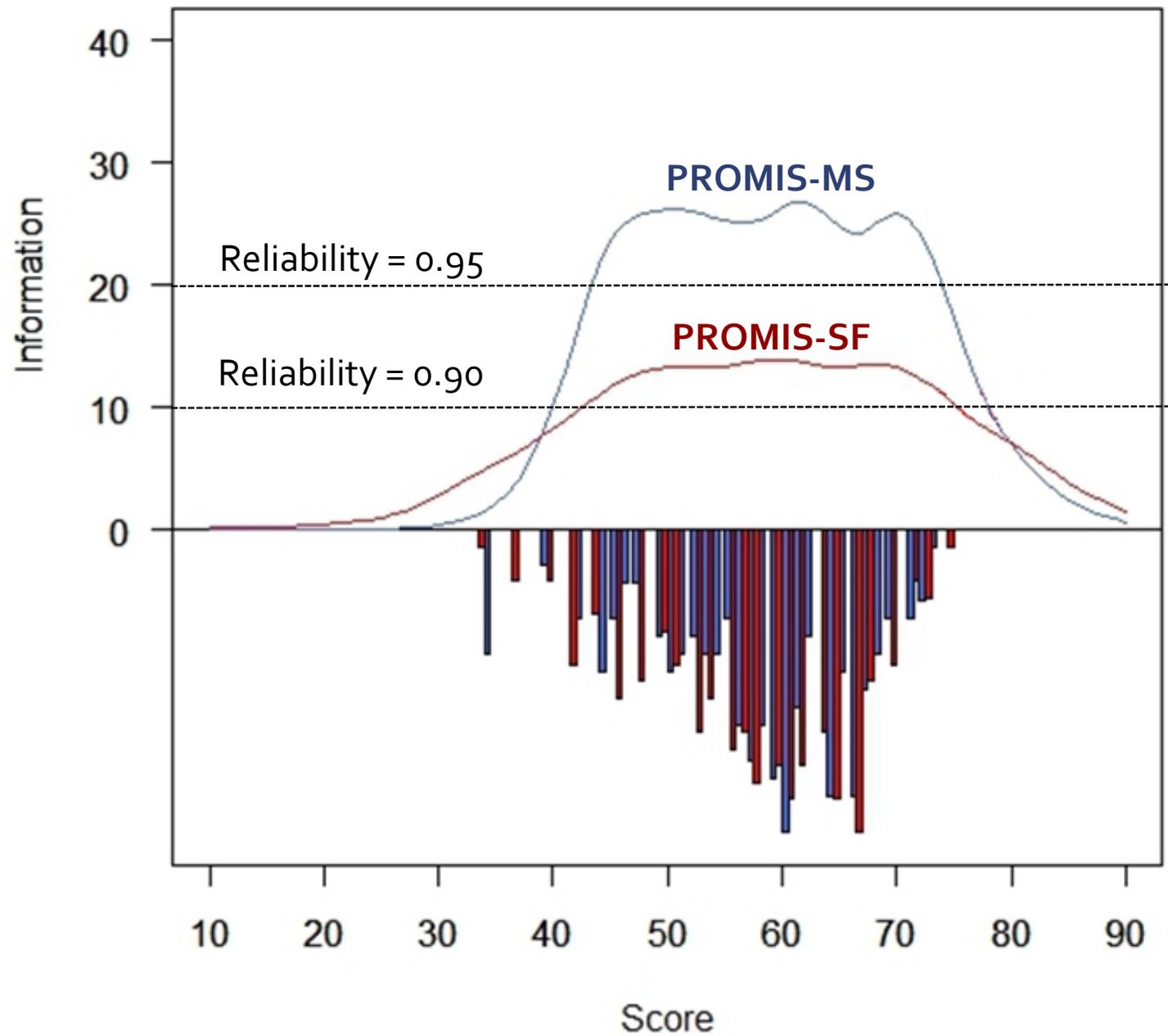
Customizing PROMIS Fatigue Assessment for Multiple Sclerosis:



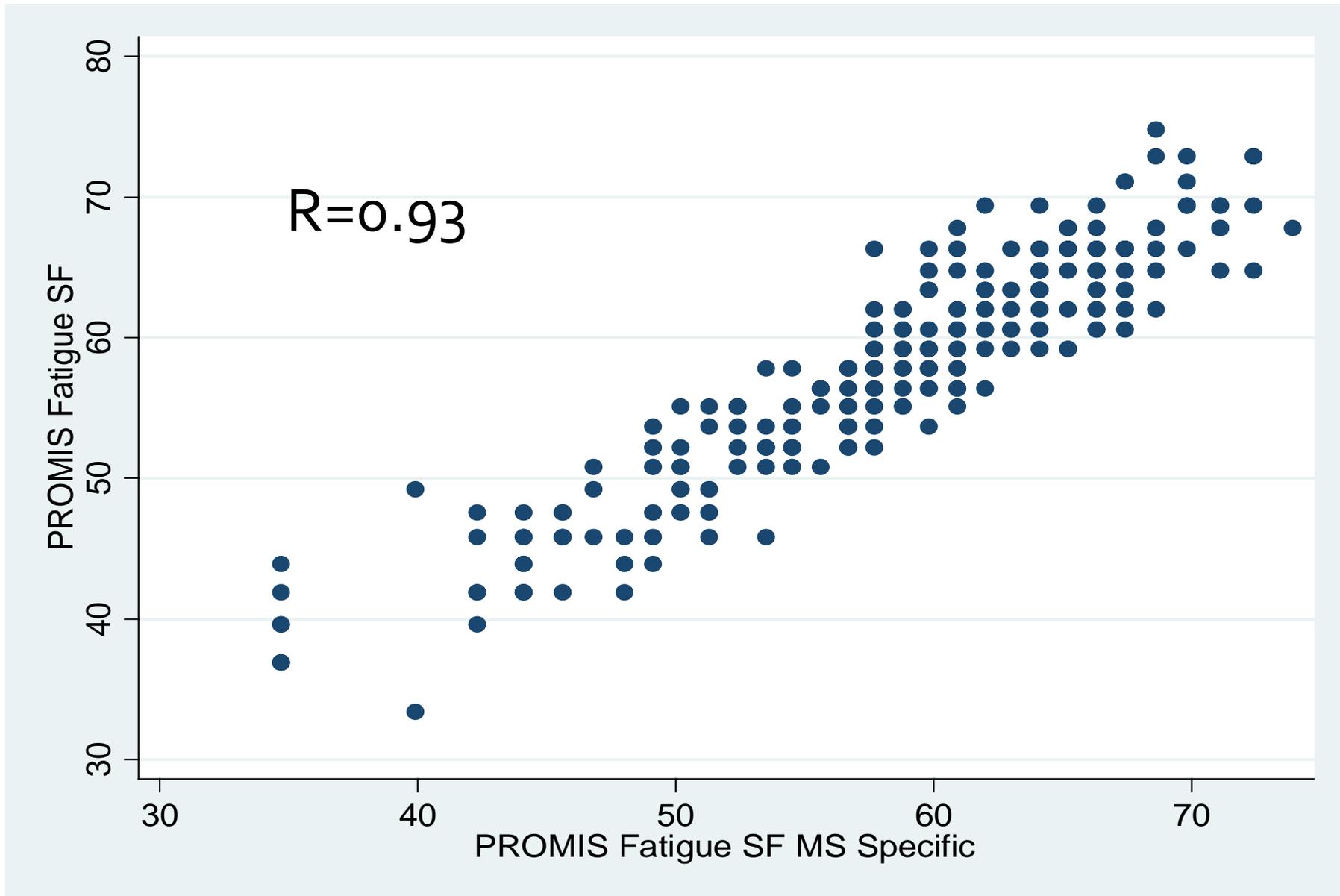
- Ranking exercise with persons with MS and expert clinicians
 - Identified 8 PROMIS fatigue items deemed most relevant to MS fatigue; created SF
 - Compared scores to off-the-shelf PROMIS 7-item generic form

PROMIS-MS	PROMIS-SF
How often were you too tired to think clearly?	How often were you too tired to think clearly?
How often did your fatigue limit you at work (include work at home)?	How often did your fatigue limit you at work (include work at home)?
How often were you too tired to enjoy life?	How often did you feel tired?
How often did you feel tired even when you hadn't done anything?	How often did you experience extreme exhaustion?
How often did you have to push yourself to get things done because of your fatigue?	How often did you run out of energy?
How often did you have trouble finishing things because of your fatigue?	How often were you too tired to take a bath or shower?
To what degree did your fatigue interfere with your physical functioning?	How often did you have enough energy to exercise strenuously?
How often did your fatigue interfere with your social activities?	

PROMIS Fatigue and PROMIS Fatigue-MS



Did it make a difference in scores?



Did it make a difference in scores?

Mean scores by fatigue severity category were virtually identical

Category	Pain Severity (0 to 10)	PROMIS-SF fatigue	PROMIS-MS fatigue
none	0	41.5	41.7
mild	1-3	43.8	43.4
moderate	4-6	51.0	51.4
severe	7-10	61.7	62.2

Modifying a Qualified Measure: *Symptoms of Major Depressive Disorder Scale (SMDDDS)*

Sonya Eremenco, MA

Associate Director, Patient-Reported Outcome (PRO) Consortium

Overview



- Background of *SMDDS*
- Rationale for modification
- Modification considerations
- Proposed modification process
- Rationale for proposed study design
- Conclusion

Background of *SMDDS*



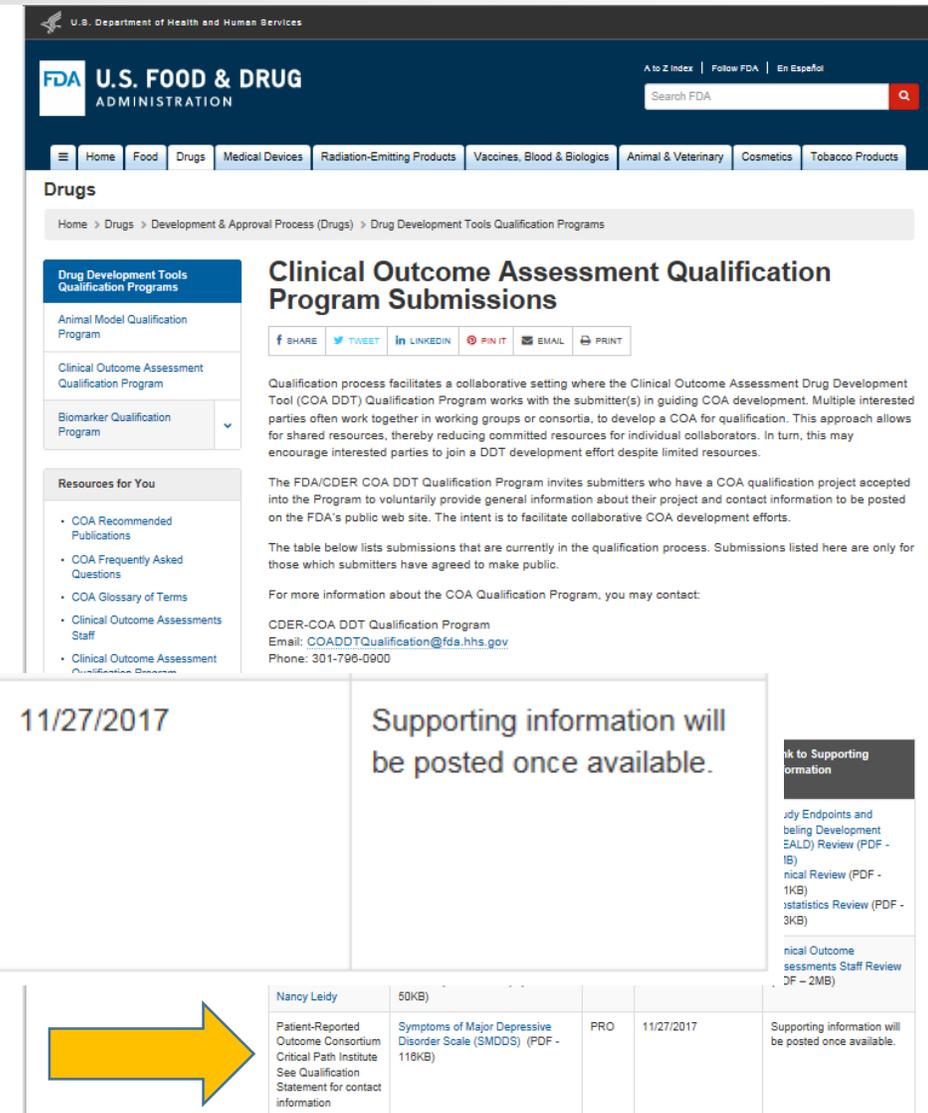
- Type of clinical outcome assessment (COA): Patient-reported outcome (PRO) instrument
- Concept of interest (COI): *SMDDS* total score measures overall symptoms of major depressive disorder (MDD)
- Context of use (COU):
 - Adults 18 years or older
 - Clinical diagnosis of MDD
 - Treated in ambulatory setting
 - Experienced major depressive episode within last 6 months
 - Hamilton Depression Rating Scale (HAM-D) score >18
 - Meets the DSM-IV or DSM-V criteria for MDD
- Attributes: intensity or frequency as a measure of severity
- Recall period: past 7 days

Conceptual Framework



Qualification

- First PRO Consortium measure to receive FDA qualification for exploratory use
- November 27, 2017
- Available for use under a licensing agreement



The screenshot shows the FDA website page for Clinical Outcome Assessment Qualification Program Submissions. The page includes a navigation menu, a search bar, and a list of qualification programs. The main content area features a section for 'Clinical Outcome Assessment Qualification Program Submissions' with a description of the qualification process and a list of resources. A table below the text lists submissions, including one for 'Symptoms of Major Depressive Disorder Scale (SMDDS)' with a date of 11/27/2017 and a status of PRO. A yellow arrow points to this submission in the table.

Patient-Reported Outcome Consortium Critical Path Institute See Qualification Statement for contact information	Symptoms of Major Depressive Disorder Scale (SMDDS) (PDF - 116KB)	PRO	11/27/2017	Supporting information will be posted once available.
				Link to Supporting Information Study Endpoints and Clinical Development EALD Review (PDF - 1B) Clinical Review (PDF - 1KB) Statistics Review (PDF - 3KB) Clinical Outcome Assessments Staff Review (PDF - 2MB)

Rationale for modification



- Request from Division of Psychiatry Products to consider modifying the *SMDDS* for MDD treatments that may have more rapid onset
 - New measurement need: detect symptom improvement in days not weeks
 - Why *SMDDS* is not adequate for newer, potentially more rapidly-acting treatments
 - 7-day recall is meant to be completed no more frequently than once per week
 - Would be unable to detect a change in symptoms in 24 hours
 - Same COI and target population: MDD symptoms in adult ambulatory population
 - Trial setting with different study design: change to COU

Modification Considerations



- Revise recall to 24 hours to create a daily diary version
- Potential revisions to content: Do items still make sense with shorter recall?
- Potential item reduction:
 - Are 16 items too many for a daily diary?
 - Should we focus on items more likely to change sooner?
 - How to identify them?
- Comparability with the 7-day recall version and need to ensure that core MDD symptoms retained

Proposed Modification Process



- Convene a new working group to sponsor the work
- C-Path and HRA teams reviewed 16-item *SMDDS* and proposed revisions
- Next steps: empirically evaluate the modifications

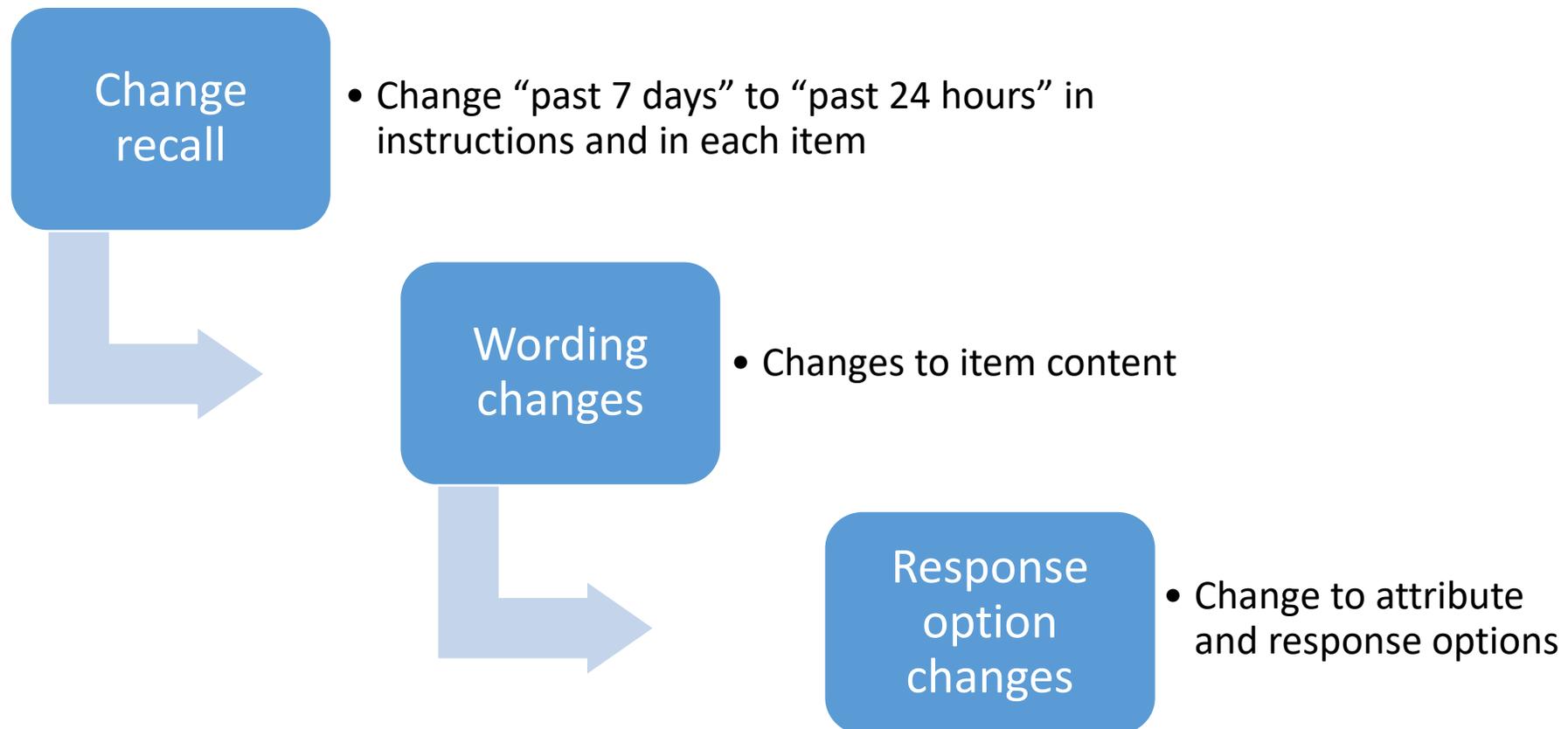
New Working Group



- Three existing PRO Consortium member firms committed to the working group in March 2018
 - Allergan
 - Boehringer Ingelheim
 - Janssen
 - Potential sponsor considering joining PRO Consortium and Working Group
- Start-up activities:
 - Project agreements with sponsors
 - Contracting with HRA as contract research organization to leverage previous qualitative and quantitative work associated with development of *SMDDS*
 - Kick-off Meeting to be scheduled

C-Path and HRA Review

- Proposed revisions



Example of items with potential issues



SMDDS (7-day recall)	SMDDS (24-hour recall)
9. Over the past 7 days, how difficult was it for you to enjoy your daily life?	9. Over the past 24 hours, how difficult was it for you to enjoy your daily life?
10. Over the past 7 days, how often did you have a problem with your sleep (falling asleep, staying asleep, or sleeping too much)?	10. Over the past 24 hours, how often how much of a problem did you have with your sleep (falling asleep, staying asleep, or sleeping too much)?
11. Over the past 7 days, how often did you have a poor appetite?	11. Over the past 24 hours, how often did you have a poor appetite?
12. Over the past 7 days, how often did you over eat?	12. Over the past 24 hours, how often did you over eat?
15. Over the past 7 days, how much of the time did you blame yourself when bad things happened?	15. Over the past 24 hours, how much of the time did you blame yourself when bad things happened?
16. Over the past 7 days, how much of the time did you feel that life is not worth living?	16. Over the past 24 hours, how much of the time did you feel that life is not worth living?

Next steps



- HRA to revisit the previous qualitative data to look for information on concepts suitable for shorter recall
- Qualitative study planned with 20 participants with MDD
 - Open-ended as well as cognitive interviewing of the revised *SMDDS* 24-hour recall
 - Confirm the revised items are understood by participants and make sense with the shorter recall period
- Letter of Intent and Qualification Plan will be prepared and submitted to FDA
- *SMDDS* 24-hour recall version will be available for Depression Working Group sponsors to use in clinical trials to collect data for evaluation of measurement properties especially ability to detect change

Rationale for proposed study design



- Qualitative vs. quantitative study
- Observational quantitative study design is not suitable to assess ability to detect change in short period of time, requires a treatment or intervention known to be rapid-acting
- Ethical issues with quantitative study with patients receiving treatment in emergency setting where quick reduction of suicidality is essential
- Cross-sectional measurement properties have been established for *SMDDS* and further quantitative work outside of a clinical trial would not provide any more information than we already have
- Qualitative study design will best answer the questions of whether all of the items are still relevant with a 24-hour recall period and whether participants understand them as intended with the revisions made

Conclusion



- *SMDDS* 7-day recall version has good cross-sectional measurement properties and provides a good foundation for modification
- New and potentially more rapid-acting treatments are driving the need for modification
- Nature of change in recall period requires qualitative data to evaluate respondent's interpretation of the items with revised recall and other changes
- Further quantitative evidence to be generated requires the appropriate setting which is not currently available outside of clinical trials
- The proposed approach is most efficient use of time and resources to meet the urgent need for a 24-hour recall version of the *SMDDS*

Panel Discussion and Q & A



Moderator

- *Maria Mattera, MPH* – Assistant Director, Patient-Reported Outcome Consortium, C-Path

Presenters

- *Elizabeth (Nicki) Bush, MHS* – Director, Patient-Focused Outcomes Center of Expertise, Eli Lilly and Company, and Industry Co-Director, PRO Consortium
- *Elektra Papadopoulos, MD, MPH* – Associate Director, COA Staff, OND, CDER, FDA
- *Dave Cella, PhD* – Professor and Chair, Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University
- *Sonya Eremenco, MA* – Associate Director, Patient-Reported Outcome Consortium, C-Path

Panelists

- *Billy Dunn, MD* – Director, Division of Neurology Products, OND, CDER, FDA