WORKSHOP ON

CLINICAL OUTCOME ASSESSMENTS (COAs)

IN CANCER CLINICAL TRIALS

April 26, 2016  ■ Silver Spring, MD

Co-sponsored by

FDA

CRITICAL PATH INSTITUTE
Session 3
Existing Options for Assessing Patient-Reported Physical Function

WORKSHOP ON CLINICAL OUTCOME ASSESSMENTS (COAs) IN CANCER CLINICAL TRIALS

April 26, 2016  ■  Silver Spring, MD

Co-sponsored by
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 Participants

Chair
- Paul G. Kluetz, MD – FDA

Presenters
- Ethan Basch, MD – University of North Carolina
- Selena Daniels, PharmD, MS – FDA
- Mogens Groenvold, MD, PhD, DSci – University of Copenhagen
- David Cella, PhD – Northwestern University Feinberg School of Medicine

Panelists
- Wen-Hung Chen, PhD – FDA
- Daniel O’Connor, MD – MHRA, EMA
- Ashley Wilder Smith, PhD, MPH – NCI
The Value of Physical Function Assessment in Cancer Clinical Trials: A Clinician Perspective

Ethan Basch, MD
April 2016
Why Clinicians Care about Physical Function

- Ability to function and carry out ADLs can be substantially impacted by cancer, and by cancer therapies.
- In advanced cancer setting, when treatment is palliative, physical functioning and symptoms are a focus of goals of care.
- Functional impairment is a major driver of ER visits, hospitalization, missed treatment doses.
- We make treatment decisions based on physical functioning:
  - Treatment choices; Dose modifications and holds; Setting goals of care.
- Therefore, a mainstay of cancer care is to understand how our patients are functioning, and how treatments will affect their functioning.
Why Clinicians Need Physical Function Information from Trials and Drug Labels

- We need to be able to explain to patients how a treatment will affect them
  - Improve and/or worsen their functioning

- We need to understand if a treatment is appropriate for a patient, given their performance status

- We need to balance benefits:harms
### Common Performance Status Measures in Oncology

<table>
<thead>
<tr>
<th>ECOG PERFORMANCE STATUS</th>
<th>KARNOFSKY PERFORMANCE STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0—Fully active, able to carry on all pre-disease performance without restriction</td>
<td>100—Normal, no complaints; no evidence of disease</td>
</tr>
<tr>
<td>1—Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work</td>
<td>90—Able to carry on normal activity; minor signs or symptoms of disease</td>
</tr>
<tr>
<td>2—Ambulatory and capable of all selfcare but unable to carry out any work activities; up and about more than 50% of waking hours</td>
<td>80—Normal activity with effort, some signs or symptoms of disease</td>
</tr>
<tr>
<td>3—Capable of only limited selfcare; confined to bed or chair more than 50% of waking hours</td>
<td>70—Cares for self but unable to carry on normal activity or to do active work</td>
</tr>
<tr>
<td>4—Completely disabled; cannot carry on any selfcare; totally confined to bed or chair</td>
<td>60—Requires occasional assistance but is able to care for most of personal needs</td>
</tr>
<tr>
<td>5—Dead</td>
<td>50—Requires considerable assistance and frequent medical care</td>
</tr>
<tr>
<td>0—Dead</td>
<td>40—Disabled; requires special care and assistance</td>
</tr>
<tr>
<td>10—Moribund</td>
<td>30—Severely disabled; hospitalization is indicated although death not imminent</td>
</tr>
<tr>
<td></td>
<td>20—Very ill; hospitalization and active supportive care necessary</td>
</tr>
</tbody>
</table>
Table 1. Baseline Characteristics of the Patients.*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cetuximab plus Platinum–Fluorouracil (N = 222)</th>
<th>Platinum–Fluorouracil Alone (N = 220)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex — no. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>197 (89)</td>
<td>202 (92)</td>
</tr>
<tr>
<td>Female</td>
<td>25 (11)</td>
<td>18 (8)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median age — yr</td>
<td>56</td>
<td>57</td>
</tr>
<tr>
<td>&lt;65 yr — no. (%)</td>
<td>183 (82)</td>
<td>182 (83)</td>
</tr>
<tr>
<td>≥65 yr — no. (%)</td>
<td>39 (18)</td>
<td>38 (17)</td>
</tr>
<tr>
<td>Karnofsky score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median score</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>80–90</td>
<td>80–90</td>
</tr>
<tr>
<td>&lt;80 — no. (%)</td>
<td>27 (12)</td>
<td>25 (11)</td>
</tr>
<tr>
<td>≥80 — no. (%)</td>
<td>195 (88)</td>
<td>195 (89)</td>
</tr>
<tr>
<td>Duration of disease — mo ‡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>15.5</td>
<td>15.8</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>10.3–27.0</td>
<td>9.5–33.5</td>
</tr>
<tr>
<td>Primary tumor site — no. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oropharynx</td>
<td>80 (36)</td>
<td>69 (31)</td>
</tr>
<tr>
<td>Hypopharynx</td>
<td>28 (13)</td>
<td>34 (15)</td>
</tr>
<tr>
<td>Larynx</td>
<td>59 (27)</td>
<td>52 (24)</td>
</tr>
<tr>
<td>Oral cavity</td>
<td>46 (21)</td>
<td>42 (19)</td>
</tr>
<tr>
<td>Other</td>
<td>9 (4)</td>
<td>23 (10)</td>
</tr>
<tr>
<td>Extent of disease — no. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only locoregionally recurrent</td>
<td>118 (53)</td>
<td>118 (54)</td>
</tr>
<tr>
<td>Metastatic with or without locoregional recurrence</td>
<td>104 (47)</td>
<td>102 (46)</td>
</tr>
<tr>
<td>Histologic type — no. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well differentiated</td>
<td>35 (16)</td>
<td>40 (18)</td>
</tr>
<tr>
<td>Moderately differentiated</td>
<td>93 (42)</td>
<td>101 (46)</td>
</tr>
<tr>
<td>Poorly differentiated</td>
<td>46 (21)</td>
<td>46 (21)</td>
</tr>
<tr>
<td>Not specified or missing</td>
<td>48 (22)</td>
<td>33 (15)</td>
</tr>
</tbody>
</table>

A ECOG Performance Score

- Abiraterone–prednisone, 12.3 mo
- Prednisone alone, 10.9 mo

Hazard ratio, 0.82 (95% CI, 0.71–0.94)

P = 0.005

No. at Risk

<table>
<thead>
<tr>
<th>Patients without Decline in ECOG Score (%)</th>
<th>Months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>33</td>
</tr>
</tbody>
</table>

NEJM, 2013 (PMID: 23228172)
Conclusion

- Clinicians/investigators value physical functioning ("performance status") information, use it to direct clinical care, to determine eligibility and stratify in clinical trials, and to evaluate treatment benefits.

- This information is often reported by clinicians on behalf of patients, not directly by patients.
Does Clinician Reporting Agree with Patients?

- Clinical investigators in a phase II lung cancer trial assigned either to see patient self-reported KPS or not
- Agreement between clinician and patient KPS measured

<table>
<thead>
<tr>
<th>Investigators who did not see PRO:</th>
<th>Patient Graded Worse</th>
<th>Agreement</th>
<th>Clinician Graded Worse</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>39%</td>
<td>22%</td>
<td>14%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Investigators who saw PRO:</th>
<th>Patient Graded Worse</th>
<th>Agreement</th>
<th>Clinician Graded Worse</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8%</td>
<td>85%</td>
<td>7%</td>
</tr>
</tbody>
</table>

Clinical Trials, 2015 (PMID: 26542025)
Conclusion

- Patients know best, and we clinicians know it.
- Clinicians often underestimate patients’ physical function impairment.
What About in Clinical Trials?

Median time to clinician-reported ECOG PS decline:
12.3 vs. 10.9 months (P=0.005)

NEJM, 2013 (PMID: 23228172)

Median time to PRO PCS decline:
11.1 vs. 5.8 months (P<0.0001)

Lancet Oncol 2013 (PMID: 24075621)
# Metrics that Make Sense To Me (as a simple clinician)

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Item Stem</th>
<th>Responses</th>
</tr>
</thead>
</table>
| PROMIS Global06     | To what extent are you able to carry out your everyday physical activities such as walking, climbing stairs, carrying groceries, or moving a chair? | 1=Not At All  
2=A Little  
3=Moderately  
4=Mostly  
5=Completely |
Metrics that Make Sense To Me (as a simple clinician)

Variable Name
PROMIS Global06

Responses
1=Not At All
2=A Little
3=Moderately
4=Mostly
5=Completely
Metrics that Make Sense To Me (as a simple clinician)

Used in 66% of published cancer trials that assess patient-reported physical function (Unpublished systematic review, courtesy Thomas Atkinson, PhD)

1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?  
   - EORTC QLQ-C30: 1 2 3 4

2. Do you have any trouble taking a long walk?  
   - EORTC QLQ-C30: 1 2 3 4

3. Do you have any trouble taking a short walk outside of the house?  
   - EORTC QLQ-C30: 1 2 3 4

4. Do you need to stay in bed or a chair during the day?  
   - EORTC QLQ-C30: 1 2 3 4

5. Do you need help with eating, dressing, washing yourself or using the toilet?  
   - EORTC QLQ-C30: 1 2 3 4
Metrics that Make Sense To Me (as a simple clinician)

<table>
<thead>
<tr>
<th>PROMIS Global-10 Response Scale: 1 – 5 1 = Poor; 5 = Excellent</th>
<th>Within-patient difference in steps per unit increase in item score (change from prior week)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain on average</td>
<td>712</td>
</tr>
<tr>
<td>Fatigue on average</td>
<td>788</td>
</tr>
<tr>
<td>Carry out physical activities</td>
<td>618</td>
</tr>
<tr>
<td>Carry out social activities</td>
<td>705</td>
</tr>
<tr>
<td>Global physical health (T-score)</td>
<td>101</td>
</tr>
</tbody>
</table>

Qual Life Res, 2016 (PMID: 26577763)
Final Conclusions

- Physical function is a clinically important, meaningful, actionable outcome in oncology that directs treatment decisions
- Clinicians are familiar with this concept
- Patients are in the best position to provide this information
- Clinicians and patients need this information to make informed treatment decisions
- It should be measured as a PRO and reported for all trials in advanced cancer
Patient-Reported Physical Function

Selena R. Daniels, Pharm.D., M.S.
Clinical Outcome Assessments Staff (formerly SEALD)
Office of New Drugs
Center for Drug Evaluation and Research
U.S. Food and Drug Administration
Disclaimer

The views expressed in this presentation are those of the speaker, and do not necessarily represent an official FDA position.
Rationale and Focus for Patient-Reported Physical Function

• FDA is exploring the following core concepts in cancer clinical trials that are considered more directly related to treatment effect and can satisfy our regulatory requirements for well-defined assessments
  – Disease-related Symptoms*
  – Treatment-related Symptoms*
  – Physical function
  *We acknowledge separation of some these concepts will be challenging

• Patients have reported that maintaining their ability to carry out their activities and care for themselves is important

• Patient-reported physical function (PF) referred to as:
  – A person’s assessment of his/her ability to carry out important and meaningful day-to-day activities (e.g., self care, domestic) that require physical effort

Evidentiary Standards to Document Treatment Benefit & Labeling Claims

• Documented by “Substantial evidence” (21 CFR 201.56(a)(3))

• Evidence from “Adequate and well-controlled clinical trials”

• The methods of assessment are “well-defined and reliable” (21 CFR 314.126)
Instruments Highlighted

• FDA is currently evaluating existing patient-reported physical function instruments for use in cancer clinical trials

• Panel Session Focus:
  – EORTC QLQ-C30 Physical Functioning domain and item bank
  – PROMIS® Physical Function item bank

Well-Defined
Customizable
Computerized Adaptive Testing (CAT) Capability
Next Steps

• FDA will continue to explore the use of existing patient-reported PF instruments from a regulatory standpoint
  – PROMIS® Physical Function Item Bank has been submitted to the FDA Clinical Outcome Assessment Drug Development Tool Qualification Program

• FDA will continue to evaluate different analysis and data presentation methods to best address different clinical trial objectives
  – Goal: Standard analysis methods that provide the most informative, responsive, and useful data on patient-reported physical function across advanced/metastatic cancer trials
Good Measurement Principles

- Well-Defined
- Reliable
- Valid
- Detects Change
- Interpretable
Acknowledgements

- Physical Function Working Group
  - Wen-Hung Chen (FDA/CDER/OND/IO COA Staff)
  - Selena Daniels (FDA/CDER/OND/IO COA Staff)
  - Paul Kluetz (FDA/CDER/OND OHOP)
  - Laura Lee Johnson (FDA/CDER/OTS/OB)
  - Marian Strazzeri (FDA/CDER/OTS/OB)

- COA Staff
Assessing Patient-Reported PF (and RF?) with the EORTC QLQ-C30 and EORTC CAT Item Bank

Mogens Groenvold, MD, PhD, DMSc
Professor of Palliative Care and Quality of Life Assessment
University of Copenhagen and Bispebjerg Hospital
Copenhagen, Denmark
On behalf of the EORTC Quality of Life Group
Outline

• EORTC QLQ-C30 – a static questionnaire
  • EORTC QLQ-C30 Physical and Role Function scales

• EORTC QLQ-CAT
  • EORTC QLQ-CAT Physical and Role Function item banks
  • EORTC QLQ-CAT Physical and Role Function short-forms

• Comparison of content:
  • PROMIS Physical Function vs. EORTC QLQ-CAT Physical and Role Function item banks

• Conclusions
  • EORTC QLQ-C30 is widely used and accepted
  • PROMIS and EORTC PF items banks quite similar and can be linked
The EORTC QLQ-C30

• Most widely used PROM in oncology (PubMed ‘hits’ N>2,500)
• Available in 98 languages, linguistically validated
• Extensive international validation, information on interpretation (MID), etc.
• Reference values, general population samples
• Cross-cultural validation through DIF analysis in > 10,000 patients in 20 countries showed cultural equivalence
• Prognostic significance
• Used in clinical practice
• Utility measure (EORTC QLU-10D) in development
The EORTC QLQ-C30: structure and content

30 items assessing 15 domains

**Functions**
- 5 Physical function (F)
- 2 Role F
- 4 Emotional F
- 2 Social F
- 2 Cognitive F
- 2 Global health status/QoL

**Symptoms**
- 3 Fatigue
- 2 Pain
- 2 Nausea/vomiting
- 1 Dyspnea
- 1 Insomnia
- 1 Appetite loss
- 1 Constipation
- 1 Diarrhea
- 1 Financial difficulties
EORTC QLQ-C30 PF scale

1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?
2. Do you have any trouble taking a long walk?
3. Do you have any trouble taking a short walk outside of the house?
4. Do you need to stay in bed or a chair during the day?
5. Do you need help with eating, dressing, washing yourself or using the toilet?

Not at All     A Little     Quite a Bit     Very Much
EORTC QLQ-C30 PF scale

1. Do you have any trouble doing **strenuous activities**, like carrying a heavy shopping bag or a suitcase?
2. Do you have any trouble taking a **long walk**?
3. Do you have any trouble taking a **short walk** outside of the house?
4. Do you need to **stay in bed or a chair** during the day?
5. Do you **need help** with eating, dressing, washing yourself or using the toilet?

Not at All    A Little    Quite a Bit    Very Much
EORTC QLQ-C30 Role F scale

During the past week:

6. Were you limited in doing either your work or other daily activities?
7. Were you limited in pursuing your hobbies or other leisure time activities?

Not at All    A Little    Quite a Bit    Very Much
EORTC QLQ-C30 Role F scale

During the past week:
6. Were you limited in doing either your work or other daily activities?
7. Were you limited in pursuing your hobbies or other leisure time activities?

Not at All  A Little  Quite a Bit  Very Much
EORTC CAT project - aims

- To move from a static questionnaire to an adaptive questionnaire by
  - Developing item banks for CAT for each of the 14 dimensions of the EORTC QLQ-C30 (except overall health/QL)
- Methodology largely similar to PROMIS
EORTC ‘rules’ for item bank development

• For each HRQOL dimension, the new instrument should:
  • Measure the **same aspects** (sub-dimensions) as the QLQ-C30 scale
  • **Include the legacy** (QLQ-C30) items
  • Be **backward compatible** with the QLQ-C30
  • Use **similar “item style”** as the QLQ-C30 items (same response options, timeframe etc.)

(This approach is different from most others, such as PROMIS, due to our insistence on reproducing an existing instrument)
Procedure for item bank development

• **Phase I**: Literature search
  • Define concept and identify items in existing instruments

• **Phase II**: Formulation of new items and expert evaluations
  • Item list **trimmed** to non-redundant items fitting the definition
  • Based on this list **new items formulated**
  • Items **evaluated by international experts**
**EORTC CAT item bank development**

- **Phase III**: Pre-testing
  - Interviews with international sample of cancer patients (min. 3*10 patients)

- **Phase IV**: Field-testing and psychometric analyses
  - Obtain responses from about **1,000 cancer patients** (international, mixed sample)
  - **Psychometric analyses** including
    - Factor analysis to evaluate dimensionality
    - Calibration and evaluation of IRT model
    - Evaluation of DIF
Development of computerised adaptive testing (CAT) for the EORTC QLQ-C30 dimensions – General approach and initial results for physical functioning

Morten Aa. Petersen a,*, Mogens Groenvold a,b, Neil K. Aaronson c, Wei-Chu Chie d, Thierry Conroy e, Anna Costantini f, Peter Fayers g,h, Jorunn Helbostad i, Bernhard Holzner j, Stein Kaasa k, Susanne Singer l, Galina Velikova m, Teresa Young n, on behalf of the EORTC Quality of Life Group
Development of EORTC PF Item Bank (31 items)

**Step 1:** 975 PF items identified. 407 items deemed relevant. Many redundant.

**Step 2:** 86 new items constructed.

**Step 3:** Reduced to 51 new items + 5 QLQ-C30 PF items

**Step 4:** Collected 1,176 patient responses to the 56 items from six countries (DK, FR, GE, IT, TW, UK).

**Psychometric/IRT analyses**
- Factor analyses: 34 items could be include in 1-dimensional model: RMSEA=0.09, CFI=0.94, TLI=0.98
- IRT calibration: 31 items had acceptable fit to IRT model: all item residuals ≤ 0.01, all item RMSEs ≤ 0.68
- DIF particularly across countries. Had little impact on PF estimation.
- **Hence, 31 items in PF item bank.**
## PF in EORTC CAT vs. PROMIS item banks

### EORTC (ICF) (N=31)
- Self-care
- Lifting and carrying objects
- Walking and moving
- Mobility/unspecified

### PROMIS (N=124)
- IADL
- Back and neck (central)
- Upper extremity
- Mobility/lower extremity

**Conclusion:** Quite similar, mostly overlapping, PROMIS a bit broader

- Rose M et al., JCE 2014
- Petersen MA et al. EJC 2010
The EORTC PF Short-Form for advanced cancer: the 5 original FF items + 4 new = 9 items (draft)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Do you have any trouble walking for 30 min.?</td>
</tr>
<tr>
<td>7</td>
<td>Do you have any trouble hiking 3 km on uneven surfaces?</td>
</tr>
<tr>
<td>8</td>
<td>Do you have any trouble carrying a heavy bag upstairs?</td>
</tr>
<tr>
<td>9</td>
<td>Do you have any trouble running a short distance, such as to catch the bus?</td>
</tr>
</tbody>
</table>
Information curves compared to score distribution for advanced cancer patients: the original 5 item PF scale (IRT scored) (left) and 9 item PF short-form (right)

The 9-item short-form gives a 2.5 fold increase in information and a better match between the instrument and the target population
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12.</td>
<td>Have you needed assistance in doing your work or daily activities?</td>
</tr>
<tr>
<td>13.</td>
<td>Have you been limited in doing light housework (e.g. dusting or making the bed)?</td>
</tr>
<tr>
<td>14.</td>
<td>Have you been limited in doing heavy housework (e.g., washing floors or vacuuming)?*</td>
</tr>
<tr>
<td>15.</td>
<td>Have you been limited in doing physically demanding recreational activities (e.g., swimming or cycling)?</td>
</tr>
</tbody>
</table>

Some of these items are similar to PROMIS PF items.
Summary/conclusions

• EORTC PF scale is **extensively used, translated, and validated** (in all ways) for oncology trials

• **Thousands of published studies** available for comparison – maximal clinical interpretability (PubMed N>2,500)

• **Available as the 5-item QLQ-C30 scale** or as **EORTC CAT**, e.g., as a 9-item PF short-form (including the original 5 items): produces original sum score as well as IRT score

• **EORTC RF (role function): a related but distinct** domain – can be assessed with the 2-item QLQ-C30 scale or as EORTC CAT, e.g., as a 6-item PF short-form (including the original 2 items)

• Some PROMIS PF items could be categorized as EORTC RF

• Good opportunities for **cross-walk/calibration** between PROMIS and EORTC PF
Thank you to...

- FDA and C-Path for invitation to present
- EORTC Quality of Life Group
- EORTC Quality of Life Group CAT project group
  - Lead statistician Morten Aagaard Petersen
Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust.
The PROMIS® Physical Function Item Bank and a Unifying Physical Function Metric

David Cella, PhD
Northwestern University

www.healthmeasures.net
Overview

• Development of the PROMIS PF Item Bank

• PROMIS as the basis for a common (unifying) PF metric

• Case study: FACT-Physical Well-Being & PROMIS PF
The Patient Reported Outcomes Measurement Information System (PROMIS) Domain Framework

Self-Reported Health
- Global Health
  - Physical Health
    - Symptoms
    - Function
  - Mental Health
    - Affect
    - Behavior
    - Cognition
  - Social Health
    - Relationships
    - Function
Definition and Scope of Physical Function

• “Ability to carry out various activities that require physical capability, ranging from self-care (activities of daily living) to more vigorous activities.”

• “Universal,” not disease-specific

• Subdomains include
  • Mobility
  • Upper extremity function
  • Axial (neck and back) function

(Bruce, Fries, Ambrosini, Lingala, Gandek, Rose, & Ware, 2009)
PROMIS 1 (2004-2009)
Development of the Universal PF Bank

• Instrument review: 1,728 items from 168 instruments

• Binning/Winnowing of items

• Focus groups, cognitive interviews, surveys on clarity of representative items (RA/OA patients; older adults)

• Field testing of 168 items (N = 16,000)

(Bruce et al., 2009; Rose et al., 2014)
PROMIS Cancer Supplement (2006-2009)

• Overlapped and followed PROMIS 1 development
  • (Garcia et al., 2007)

• Qualitative work with cancer patients (n = 21; n = 40)
  • PROMIS general PF items encompassed majority of identified concepts

• Field testing (N = 521)
  • Tested for DIF between cancer and general population
  • Most cancer PF items (83%) are in universal bank

• Cancer-relevant short form development
  • Content validity documented
  • All items in general 121 item PF item bank, with common calibrations
  • ID established as 4-6 points (Yost, Eton, Garcia, & Cella, 2011)
**PROMIS® Short Form v1.2 Physical Function 10b**

<table>
<thead>
<tr>
<th>Question</th>
<th>Rating Options</th>
</tr>
</thead>
</table>
| Does your health now limit you in doing vigorous activities, such as running, lifting heavy objects, or participating in strenuous sports? | 5 = Not at all  
4 = Very little  
3 = Somewhat  
2 = Quite a lot  
1 = Cannot do |
| Are you able to do chores such as vacuuming or yard work?                 | 5 = Without any difficulty  
4 = With a little difficulty  
3 = With some difficulty  
2 = With much difficulty  
1 = Unable to do |
| Are you able to get in and out of a car?                                 | 5 = Without any difficulty  
4 = With a little difficulty  
3 = With some difficulty  
2 = With much difficulty  
1 = Unable to do |
| Are you able to go up and down stairs at a normal pace?                  | 5 = Without any difficulty  
4 = With a little difficulty  
3 = With some difficulty  
2 = With much difficulty  
1 = Unable to do |
| Are you able to run errands and shop?                                    | 5 = Without any difficulty  
4 = With a little difficulty  
3 = With some difficulty  
2 = With much difficulty  
1 = Unable to do |
| Does your health now limit you in bathing or dressing yourself?          | 5 = Not at all  
4 = Very little  
3 = Somewhat  
2 = Quite a lot  
1 = Cannot do |
| Are you able to bend down and pick up clothing from the floor?           | 5 = Without any difficulty  
4 = With a little difficulty  
3 = With some difficulty  
2 = With much difficulty  
1 = Unable to do |
| Are you able to lift 10 pounds (5 kg) above your shoulder?                | 5 = Not at all  
4 = Very little  
3 = Somewhat  
2 = Quite a lot  
1 = Cannot do |
| Does your health now limit you in putting a trash bag outside?           | 5 = Not at all  
4 = Very little  
3 = Somewhat  
2 = Quite a lot  
1 = Cannot do |
| Does your health now limit you in doing moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf? | 5 = Not at all  
4 = Very little  
3 = Somewhat  
2 = Quite a lot  
1 = Cannot do |
### CaPS and PROMIS-2 Testing

<table>
<thead>
<tr>
<th></th>
<th>Jensen et al., 2015 (N=4,840)</th>
<th>Yost et al., 2011 (N=310)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PF16</td>
<td>PF4a</td>
</tr>
<tr>
<td><strong>PROMIS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>0.76</td>
<td>0.74</td>
</tr>
<tr>
<td>Fatigue</td>
<td>-0.72</td>
<td>-0.67</td>
</tr>
<tr>
<td>Pain Interference</td>
<td>-0.67</td>
<td>-0.64</td>
</tr>
<tr>
<td>Depression</td>
<td>-0.50</td>
<td>-0.47</td>
</tr>
<tr>
<td>Anxiety</td>
<td>-0.48</td>
<td>-0.45</td>
</tr>
<tr>
<td>Sleep Disturbance</td>
<td>-0.41</td>
<td>-0.38</td>
</tr>
<tr>
<td><strong>FACT PWB</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global General Health</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Global Physical Health</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Global Physical Functioning</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Global Fatigue</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>FACIT-Fatigue</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SF-36 PF10</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Global Physical Limitations (CaPS)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>ECOG PSR</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Establishing a common metric for PROMIS Physical Function
PROsetta Stone Linking Project

• A known impediment to CER: myriad of instruments available to measure same construct (Sox & Greenfield, 2009)

• Based in educational testing, “linking” establishes a way to equate scores across measures (Kolen & Brennan, 2004)

• PROsetta Stone project established 53 links with PROMIS/Toolbox

• Multi-method approach, including item response theory (Kolen & Brennan, 2004; Choi et al., 2014)
PROsetta Stone Example: SF-36 PF

• SF-36 v2 Physical Function
  • 10-item PF-10

• PROMIS Physical Function
  • 76 items served as anchor for linking

• Co-administered both instruments in PROMIS 1 (N = 733)
• Instruments highly correlated (r = .89-.91)
• Estimated SF-36 PF parameters on the PROMIS metric

(Schalet, Revicki, Cook, Krishnan, Fries, & Cella, 2015)
SF-36 PF-10 and HAQ (Health Assessment Questionnaire) on PROMIS PF metric

(Schalet, Revicki, Cook, Krishnan, Fries, & Cella, 2015)
## PROsetta Stone:
### SF-36 PF-10 to PROMIS

<table>
<thead>
<tr>
<th>SF-36 PF-10 Score</th>
<th>PROMIS T-score</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>24.5</td>
<td>4.0</td>
</tr>
<tr>
<td>11</td>
<td>28.3</td>
<td>2.8</td>
</tr>
<tr>
<td>12</td>
<td>30.3</td>
<td>2.5</td>
</tr>
<tr>
<td>13</td>
<td>32.0</td>
<td>2.2</td>
</tr>
<tr>
<td>14</td>
<td>33.4</td>
<td>2.1</td>
</tr>
<tr>
<td>15</td>
<td>34.8</td>
<td>2.0</td>
</tr>
<tr>
<td>16</td>
<td>36.0</td>
<td>2.0</td>
</tr>
<tr>
<td>17</td>
<td>37.2</td>
<td>2.0</td>
</tr>
<tr>
<td>18</td>
<td>38.4</td>
<td>1.9</td>
</tr>
<tr>
<td>19</td>
<td>39.5</td>
<td>1.9</td>
</tr>
<tr>
<td>20</td>
<td>40.7</td>
<td>1.9</td>
</tr>
<tr>
<td>21</td>
<td>41.8</td>
<td>1.9</td>
</tr>
<tr>
<td>22</td>
<td>42.9</td>
<td>1.9</td>
</tr>
<tr>
<td>23</td>
<td>44.1</td>
<td>2.0</td>
</tr>
<tr>
<td>24</td>
<td>45.3</td>
<td>2.0</td>
</tr>
<tr>
<td>25</td>
<td>46.7</td>
<td>2.1</td>
</tr>
<tr>
<td>26</td>
<td>48.2</td>
<td>2.3</td>
</tr>
<tr>
<td>27</td>
<td>49.9</td>
<td>2.5</td>
</tr>
<tr>
<td>28</td>
<td>52.0</td>
<td>2.9</td>
</tr>
<tr>
<td>29</td>
<td>55.0</td>
<td>3.5</td>
</tr>
<tr>
<td>30</td>
<td>61.7</td>
<td>5.7</td>
</tr>
</tbody>
</table>

[www.prosettastone.org](http://www.prosettastone.org)
Linking EORTC QLQ-30 to PROMIS metric

- Item content very similar to both PROMIS and SF-36 items
- Likely a robust link

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Not at All</th>
<th>A Little</th>
<th>Quite a Bit</th>
<th>Very Much</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2.</td>
<td>Do you have any trouble taking a long walk?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3.</td>
<td>Do you have any trouble taking a short walk outside of the house?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4.</td>
<td>Do you need to stay in bed or a chair during the day?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5.</td>
<td>Do you need help with eating, dressing, washing yourself or using the toilet?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
FACT-Physical Well-Being & PROMIS Physical Function

Or: “What to do when you have a good scale that doesn’t align perfectly with a conceptual definition of physical function”
Measuring Your Health (MY-Health) Study

• A PROMIS-2 study (Potosky, Jensen, Moinpour)

• Adult cancer patients recruited through 4 SEER cancer registries in three states (CA, LA, NJ) between 2011-2013

• Diagnosed with 7 cancers (breast, prostate colorectal, non-small cell lung, Non-Hodgkin lymphoma, uterine, or cervical cancer)

• Co-administered FACT-G PWB and PROMIS 16-item PF SF
PROMIS Physical Functioning & FACT-G Physical Well-Being

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Baseline Mean</th>
<th>Baseline SD</th>
<th>Follow-up Mean</th>
<th>Follow-up SD</th>
<th>Change</th>
<th>Overall Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROMIS PF</td>
<td>2958</td>
<td>46.0</td>
<td>9.7</td>
<td>46.8</td>
<td>10.1</td>
<td>0.8</td>
<td>0.08</td>
</tr>
<tr>
<td>FACT-G PWB</td>
<td>2941</td>
<td>22.2</td>
<td>6.0</td>
<td>22.8</td>
<td>5.8</td>
<td>0.6</td>
<td>0.10</td>
</tr>
</tbody>
</table>

- Time 1 Correlation: 0.710
- Time 2 Correlation: 0.712
## PROMIS PF Effect Sizes by FACT-G PWB Change Scores

<table>
<thead>
<tr>
<th>FACT-G PWB Change Score Category</th>
<th>N</th>
<th>PROMIS PF Mean Change (SD)</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much Worse (-5 or more)</td>
<td>175</td>
<td>-6.03 (7.30)</td>
<td>-0.70</td>
</tr>
<tr>
<td>Worse (-3 to -5)</td>
<td>266</td>
<td><strong>-2.47</strong> (5.18)</td>
<td>-0.27</td>
</tr>
<tr>
<td>No Change (-2 to 2)</td>
<td>1996</td>
<td>0.73 (5.49)</td>
<td>0.08</td>
</tr>
<tr>
<td>Improved (3 to 5)</td>
<td>205</td>
<td><strong>3.49</strong> (5.35)</td>
<td>0.42</td>
</tr>
<tr>
<td>Much Improved (5 or more)</td>
<td>292</td>
<td>6.25 (6.54)</td>
<td>0.89</td>
</tr>
</tbody>
</table>
FACT-G Physical Well Being (PWB) Items

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
FACT-G Physical Well Being (PWB) Items

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
The Best Link (if only PWB is Available) – Baseline Data

<table>
<thead>
<tr>
<th>Model</th>
<th>Correlation</th>
<th>Mean Difference</th>
<th>SD Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-item</td>
<td>.760</td>
<td>-1.235</td>
<td>7.080</td>
</tr>
<tr>
<td>5-item</td>
<td>.772</td>
<td>-1.251</td>
<td>6.767</td>
</tr>
<tr>
<td>2-item</td>
<td>.704</td>
<td>-2.197</td>
<td>7.269</td>
</tr>
</tbody>
</table>
Augmenting the 2-item PWB

• The 2-item version has content that best reflects PROMIS-PF
• But there are not enough items to adequately represent PF or cover floor and ceiling
• Solution: Augment FACT-G with 3 PROMIS PF items, creating a 5-item scale (two from FACT-G-PWB and three from PROMIS-PF)
Recommended Items

- 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?

*NOTE: PFA11 and PFA53 are also two items from the 4-item short form.*
## Augmented 5-item PWB vs PROMIS PF

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
<th>Mean Difference</th>
<th>SD Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern Scoring vs. 16-item PF</td>
<td>0.95</td>
<td>0.07</td>
<td>3.23</td>
</tr>
<tr>
<td>Pattern Scoring vs. 13-item PF</td>
<td>0.91</td>
<td>0.29</td>
<td>4.26</td>
</tr>
<tr>
<td>Sum Scoring vs. 16-item PF</td>
<td>0.93</td>
<td>-0.27</td>
<td>3.59</td>
</tr>
<tr>
<td>Sum Scoring vs. 13-item PF</td>
<td>0.90</td>
<td>-0.05</td>
<td>4.47</td>
</tr>
</tbody>
</table>
Thank you
Augmented Scores

Augmented scores continue to exhibit floor and ceiling effects compared to the general bank, but these are less than the FACT-G-PWB alone. Additionally, there is significantly less bias.
Score Comparisons

The top histogram represents pattern-based EAP scoring of the 5-item augmented FACT.

The second histogram represents sum-score EAP scoring of the 5-item augmented FACT.

The third histogram represents pattern-based EAP scoring of the full 16-item PROMIS PF dataset. This includes the three items used to augment the FACT.

The fourth histogram represents pattern-based EAP scoring of the 13-item PROMIS PF dataset, i.e. no overlapping items with the augmented FACT.
Assessing Physical Function in Cancer Clinical Trials
Session Participants

Chair
- Paul G. Kluetz, MD – FDA

Presenters
- Ethan Basch, MD – University of North Carolina
- Selena Daniels, PharmD, MS – FDA
- Mogens Groenvold, MD, PhD, DSci – University of Copenhagen
- David Cella, PhD – Northwestern University Feinberg School of Medicine

Panelists
- Wen-Hung Chen, PhD – FDA
- Daniel O’Connor, MD – MHRA, EMA
- Ashley Wilder Smith, PhD, MPH – NCI
WORKSHOP ON
CLINICAL OUTCOME ASSESSMENTS (COAs)
IN CANCER CLINICAL TRIALS

April 26, 2016 - Silver Spring, MD

Co-sponsored by:

[Logo of FDA] [Logo of Critical Path Institute]