Clinical Outcome Measures in HD: Beyond UHDRS
Glenn T. Stebbins, PhD

Kick-Off Meeting, Sheraton Silver Spring Hotel, November 6-7, 2017
Disclosures

• **Consulting and Advisory Board Membership with honoraria**: Acadia, Pharmaceuticals, Adamas Pharmaceuticals, Inc., Biogen, Ceregene, Inc., CHDI Management, Inc., Ingenix Pharmaceutical Services (i3 Research), Neurocrine Biosciences, Inc., Pfizer, Inc., Ultragenyx, Inc..

• **Grants and Research**: National Institutes of Health, Michael J. Fox Foundation for Parkinson’s Research, Dystonia Coalition, CHDI, International Parkinson and Movement Disorder Society, CBD Solutions.

• **Honoraria**: American Academy of Neurology, Biogen, Inc., Food and Drug Administration, International Parkinson and Movement Disorder Society, Michael J. Fox Foundation for Parkinson’s Research, National Institutes of Health.

• **Salary**: Rush University Medical Center
Unified Huntington’s Disease Rating Scale

- Originally published 1996 – Revised 1999
- Assesses Motor, Cognitive, Behavioral and Functional capacity
- ClinRO administration
- Good internal consistency, inter-rater reliability for Motor
- Sensitivity to change Motor, TCF and Behavior
Unified Huntington’s Disease Rating Scale

• Delphi panel – yes?
• Review of existing scales - yes
• Focus groups – no
• Cognitive pretesting patients – no
• Cognitive pretesting HCP – no
• Translations – yes, but not validated process

Goncalves JNNP 2014
Unified Huntington’s Disease Rating Scale

- Motor assessment – good clinimetrics
- Behavior, Functional and Cognitive – not so good
  - Behavioral – ClinRO, no patient input, multiplicative scoring
  - Functional – ClinRO, no patient input, scoring issues
  - Cognitive – Limited domains, no patient input, raw score scoring
Cognitive Assessment—HD-CAB

- Topic worthy of separate conference!
- Excellent review chapter in Handbook of Clinical Neurology (vol 144 – Huntington Disease) by Dr. Stout
  - Basic framework for assessing cognition in clinical trials – focus on HD
  - HD-CAB development and validation
  - Pathway for continued development of cognitive outcome in HD
Expanding Behavioral and Functional Assessments – 2 Approaches

• Modify existing scales
  • Hospital Anxiety Depression Scale – Snaith Irritability Scale

• Develop new scales
  • Problem Behavior Assessment – Short

• FuRST 1.0
Datasets

• Three datasets
  - Enroll-HD-Plus-Perodic-Dataset-2016-10-R1
    • 8,165 Registry data records (HDCAT & ID)
    • 17,187 Enroll data records (HDCAT & ID)
  - Enroll-SPS014-2016-R1
    • 2,098 Registry data records (HADS-SIS & ID)
    • 17,187 Enroll data records (HADS-SIS & ID)
  - Data-project-0844
    • 25,636 Data records
    • 7,451 Individual HADS-SIS item scores

• Combine and Clean
  - Match on ID
  - Merge with HDCAT, HADS-SIS, PBA-s, FuRST, demographics
  - Select only BASELINE visits (n = 8714) and BASELINE for HDCAT Manifest (n = 4752)
HADS-SIS

• Combine two scales with some overlap
  • Original structure – 2 factors for HADS (depression and anxiety) and 2 factors for SIS (inward and outward irritability)
  • Translated into 11 languages - ? Method

• Clinimetrics
  • Reliability acceptable
  • Construct validity
    • Factor structure different
      • two factor is a more parsimonious explanation (negative affect; agitated affect)
  • Translation process inadequate
Problem Behaviors Assessment

• **11 items shorted version of the PBA-HD**
  - Developed by “expert panel”
  - ClinRO
  - Multiplicative scoring (severity * frequency)

• **Clinimetrics**
  - Reliability low but acceptable
  - Possible “floor effect” for 8 items
  - Poor factor structure results
  - Impossible score values due to multiplicative scoring
    - 5, 7, 9, 10, 11, 13, 14, 15 impossible values – changes distribution and adversely affects variance structure
Functional Rating Scale Taskforce

FuRST 1.0

• 2010 – FuRST workgroups
  • Expert panels (Delphi group)
  • Focus groups (pre-manifest HD; early HD; caregivers) 101 in person: 16 telephone.
  • Item development
  • Field testing
  • Item refinement

• 2012 – 21 item FuRST 1.0 developed
  • distribution and adversely affects variance structure
### 1. Irritability/Anger

<table>
<thead>
<tr>
<th>Symptom Intensity</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Absent</strong></td>
<td>Never/Absent</td>
</tr>
<tr>
<td>Not irritable</td>
<td>0</td>
</tr>
<tr>
<td><strong>Mild</strong></td>
<td>1</td>
</tr>
<tr>
<td>Somewhat irritable, but no overt expression of irritability.</td>
<td></td>
</tr>
<tr>
<td><strong>Moderate</strong></td>
<td>1</td>
</tr>
<tr>
<td>Irritable, minor autonomic arousal (e.g. slight flushing or palpitations), no overt expression of irritability</td>
<td></td>
</tr>
<tr>
<td><strong>Severe</strong></td>
<td>2</td>
</tr>
<tr>
<td>Irritable, definite autonomic arousal (e.g. tremulous voice, shaking, close to tears) easily irritated and feeling angry, with mild overt expression of irritability (example, short-tempered, snap at others, expression of frustration, passive aggressive).</td>
<td></td>
</tr>
<tr>
<td><strong>Very Severe</strong></td>
<td>3</td>
</tr>
<tr>
<td>Extremely irritable, autonomic arousal with definite expression of irritability/anger (e.g. getting into arguments, shouting, cursing, loss of temper).</td>
<td></td>
</tr>
</tbody>
</table>
Limitations of FuRST 1.0

- **Construction**
  - Only telephone focus group (n = 16) data analyzed. Multilingual groups (English, Portuguese, French and Dutch) not used to identify domains.
  - Complicated GRID scoring depending on level of expertise of rater
    - No cognitive pre-testing and no follow-up Delphi review
    - No functional impact on some items

- **Field Testing (n = 96: 63 Pre-HD; 33 HD)**
  - All items ≥ 50% with scores 0 – 1
    - Floor effects?
  - Item-to-total correlations low on some items

- **Review Team recommended modification of FuRST 1.0**
Reasons for measurement failures? (avoiding mistakes of the past)

1. Definition of domain of interest
2. Development of individual items for the rating scale
3. Selection of type of scaling that is applied to items
4. Development of preliminary scale and refinement of items pool
5. Development of final rating scale with continued reliability and validity testing

- Delphi process and/or focus groups
- Content validity assessed and modifications made
- Item reduction, selection and scale validation testing
Reasons for measurement failures?

- Defining the domain of interest
  - Information on importance from the source (clinicians, patients, caregivers, other stakeholders)

- Defining assessment source (PRO, ClinRO, ObsRO, PerfO)
  - PRO (insight? FuRST 2.0 example)
  - ClinRO and ObsRO (knowledge of unknowable events)
  - PerfO (time consuming, difficult to quantify, ? relationship to disease markers)

- Defining scaling approach (summary, additive, multiplicative)

- Develop items (stem and anchor)

- Feedback from stakeholders – Cognitive Pretesting

- Modification and validation
• Wearables?

Adams – Digit Biomark 2017;1:52-63