Success in sharing data from HD natural history studies

Amrita Mohan
Director, Clinical Bioinformatics
amrita.mohan@chdifoundation.org

HD-RSC Kick-off Meeting
Silver Creek, MD
November 6, 2017
Thought structure

• Rationale for natural history (NH) studies in HD
• HD NH studies
• Evolution of study data
• HD NH data sharing process

• Data application
  – Approach
  – Select findings

• Key takeaways
HD natural history studies: Why?

Guiding clinical development

- **Specific challenges**
  1. Subtle symptoms, early on
  2. Sparse population
  3. Disease progression & progression biomarkers
HD natural history studies

• ~15 studies known so far
  – Enroll-HD largest in terms of participant pool
  – Early studies focused on clinical assessments (e.g. UHDRS)
  – More recent focusing on neuroimaging and molecular measurements
Evolution of study data

THEN & NOW

• Paper based
• Largely unstructured
• Siloed (no synthesis)
• Smaller studies/fewer measures
• Simple analysis

• EMR & Big data
• More structured
• On the fly digestion
• Multi-variable studies
• Patient privacy/de-identification
• Transparency mandates
Preparing to share HD natural history dataset

Duration: 1 – 1.5yrs
Need for data standards
Application
Can we learn about HD progression trajectories?

..covering entire HD gene expansion carrier population
Largest HD natural history Dataset

<table>
<thead>
<tr>
<th>Cohort study</th>
<th>#Approx. Participants</th>
<th>CAG</th>
<th>Max visits</th>
<th>Mean visits</th>
<th>Motor</th>
<th>Functional</th>
<th>Psychiatric</th>
<th>Cognitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enroll-HD</td>
<td>7,500</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Registry-HD</td>
<td>12,000</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Track-HD/ON</td>
<td>450</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PREDICT-HD</td>
<td>1,500</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

• Criteria
  – Large participant base
  – Longitudinal visit data
  – Clinical assessment data
Characterizing HD-NH Data

**Clinical Variables**

- Shared
  - CAG length, age, gender, BMI, education level, region, subject status, CAP, disease burden score, year to onset
- Fixed
  - UHDRS Motor, Functional, SDMT, SWRT, Trail making, among others

**Participants**

- Enroll
  - 1259
  - 37
  - 0
  - 0
  - 315
- Registry
  - 151
  - 0
  - 46
  - 57
  - 188

~2000 outcome measurements

~100 fixed measurements

~20,000 participants

**Visit statistics**

<table>
<thead>
<tr>
<th>Visit</th>
<th>Min</th>
<th>Median</th>
<th>Mean</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>17</td>
</tr>
</tbody>
</table>

**Notes**

- Shared
- Fixed
- Confidential
Leveraging HD-NH Data

- Match variables
- Merge patient records
- Unify category coding
- QC/Outlier check
- Feature selection, engineering & model development
- Missing value analysis

HD-NH Data
Observation: HDGEC population-wide trend…

\[ \text{CAP} = \text{AGE} \times \frac{\text{CAG} - 30}{6.27} \]

Using clinical outcome measures only; without CAG/CAP score
Observation:
Distinct clinical events spanning 4 decades…

- Population-wide
- Measurable using standard battery of tests
- Can be brought to clinical practice
Key takeaways

• Multi-source, heterogeneous HD clinical datasets integrated and analyzed to understand complex, population-wide phenotypes

• A clear set of HD data standards can expedite leveraging clinical trials datasets

• New knowledge can improve HD clinical practice; transform patient journey
Thank you