A Data-Driven Method for Generating Robust Symptom Onset Indicators in Huntington’s Disease Registry Data

Health Monitoring Across EHRs and Registries, Session S13

Zhaonan Sun (presenter)
Ying Li, Soumya Ghosh, Yu Cheng, Jianying Hu
IBM T. J. Watson Research Center
Amrita Mohan, Cristina Sampaio
CHDI Management/CHDI Foundation
Disclosure

I and my co-authors have no relevant relationships with commercial interests to disclose.
Disease Registry

• Collections of information about individuals with specific diseases or conditions using observational study methods
  • vs. EHR: first-hand information on a specific population
  • vs. RCT: observational data, broader research goals

• Provide health care professionals and researchers rich information for various purposes:
  • Tack the natural course of the target disease
  • Understand variations in treatment and outcomes
  • Assess clinical cost and quality of life
  • …

Characterized by the onsets and progression of symptoms
Huntington’s Disease (HD)

- Neurodegenerative genetic disorder: progressive loss of structure and function of neurons
- Symptoms: motor impairment, cognitive deterioration, behavioral disorder, functional deficits

- **Monogenic**: caused by abnormal CAG repeat expansion in the huntingtin (HTT) gene

<table>
<thead>
<tr>
<th>Repeat count</th>
<th>Classification</th>
<th>Disease status</th>
<th>Risk to offspring</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;26</td>
<td>Normal</td>
<td>Will not be affected</td>
<td>None</td>
</tr>
<tr>
<td>27–35</td>
<td>Intermediate</td>
<td>Will not be affected</td>
<td>Elevated but &lt;&lt;50%</td>
</tr>
<tr>
<td>36–39</td>
<td>Reduced Penetrance</td>
<td>May or may not be affected</td>
<td>50%</td>
</tr>
<tr>
<td>40+</td>
<td>Full Penetrance</td>
<td>Will be affected</td>
<td>50%</td>
</tr>
</tbody>
</table>

HD gene expansion carriers (HDGECs)

https://en.wikipedia.org/wiki/Huntington%27s_disease
Recent Observation Studies of HD

Enroll-HD

PREDICT-HD
An observational study of the earliest signs of Huntington disease

HiD
EUROPEAN HUNTINGTON’S DISEASE NETWORK

Cohort
COOPERATIVE HUNTINGTON'S OBSERVATIONAL RESEARCH TRIAL

TRACK-HD

TrackOnHD
Enroll-HD: example of a disease registry

Participants Profile
- Subject ID
- age
- gender
- CAG length
- education level

Comorbidity
- Subject ID
- ICD-10
- Start date
- End date

Medication
- Subject ID
- ATC code
- Start date
- End date

Other Event
- Subject ID
- Event code
- Start date
- End date

Periodic Assessments
- Subject ID
- Visit type
- Visit date
- Clinical measure 1
- Clinical measure 2

- Basic characteristics of participants, including demographics, genotype, medical history etc.
- Case group vs. control group

Clinical assessments from periodic visits, measured by clinicians or trained evaluators. Recording and monitoring symptoms, disease progression, quality of life, cost burden of the disease, etc.

Disease related information. Reported by participants or collected by evaluators. Provide additional participants characteristics.
Motivation

Method for Generating Robust Symptom Onset Indicators

- Measures can be affected by
  - Disease status
  - Non-disease-related factors → biases in measurements

- Target: better understanding of disease progression
  - The natural course of a disease can be characterized by the onset and progression of symptoms

- HD: clear definition of case and control participants
- Control group is not affected by disease progression, therefore is useful for reducing biases due to non-disease-progression-related factors
Steps (for each clinical measure)

1. Integrated HD observational data
2. Check Missing Values
3. Missing Value Imputation
4. Multiple Complete Dataset

- Case Participants
  - Apply Control Model on Case Participants
  - Prediction confidence Intervals for case participants from Control Model
  - Raw clinical measures from Case Participants

- Control Participants
  - Predictive Modeling
  - Control Model
  - Calculate Robust Clinical Measures for Case Participants

Key component: utilizes a predictive model based on control participants to adjust biases from raw clinical measurements among case participants

Bootstrap Sample

Symptom onset indicator
Generating robust symptom onset indicator

- Robust scores: distance between observed values and the PCI from control model
- Sign of robust scores: indicate the presence of a symptom
- Symptom onset indicator: change of signs in the robust scores
Integrated HD observational data

- Integrated from four large-scale observational studies: Enroll-HD, Registry, Track-HD/Track-On, Predict-HD
- Largest HD dataset to-date:
  - 16,553 HDGECs, 2716 control subjects
- Predictors in control models: demographics, medical history, study ID, region, etc.
Results: raw scores vs. robust scores

Symbol Digit Modality Test (SDMT)

• Natural aging process can affect the measures from cognitive assessments

• No clear threshold to determine the onset of cognitive symptoms

• Effects of the natural aging process can be reduced

• Signs of the robust scores indicate whether a patient has reached cognitive symptom onset or not
Results: raw scores vs. robust scores

- Education level can affect the measures from cognitive assessments
- No clear threshold to determine the onset of cognitive symptoms
- Effects of education level can be reduced
- Signs of the robust scores indicate whether a patient has reached cognitive symptom onset or not
Use case: explore natural course of HD

- Multiple clinical assessments, each target at monitoring one symptom
- Robust scores of a patient + temporal information → personalized disease progression pathway characterizing the onset of multiple symptoms
- Summarizing personalized disease progression pathways → natural course of the target disease at population level

- Patient 1
  - Chorea
  - Oculomotor
  - Bradykinesia

- Patient 2
  - Depression
  - Anxiety
  - Weight loss
  - Memory Loss

- Summarize at population level
  - Inward Irritability
  - Outward irritability

Personalized disease progression pathways
Population level disease progression pathway
Discussion

- A method for generating robust symptom onset indicator in HD observational data
  - Leverage control subjects
  - Reduce biases due to non-disease-progression-related factors
  - Signs of the robust measures have clear meaning → symptom onset indicator

- Use in disease progression modeling
  - Personalized and population level progression pathway
  - Used as an input in HD disease progression models

- Generalize to registry data of other diseases
  - HD: clear definition of control subjects
  - Other diseases: need carefully select control cohort
Data used in this work was generously provided by the research participants in ENROLL-HD, Registry, Track-HD, Predict-HD and CHDI Foundation, Inc that financially supports the study.
Thank you!

zsun@us.ibm.com