Novel tools for analysis and visualization of longitudinal data

Sudeshna Das, PhD & Deborah Blacker, MD, ScD
Massachusetts Alzheimer’s Disease Research Center
FALL ADC MEETING 2019
Novel tools for analysis and visualization of longitudinal data

Sudeshna Das, PhD & Deborah Blacker, MD, ScD
Massachusetts Alzheimer’s Disease Research Center
FALL ADC MEETING 2019
Novel Tools for analysis and visualization of longitudinal data

Sudeshna Das, PhD & Deborah Blacker, MD, ScD
Massachusetts Alzheimer’s Disease Research Center
FALL ADC MEETING 2019
Background & Objectives

• MADRC has acquired longitudinal UDS data on ~1600 participants for >10 years

• Data & Clinical Core collaborate on multiple research agenda, such as:
  • Selecting samples for specific studies
  • Understanding heterogeneity in cohort

• Case Study:
  • Develop tool to select study participants with specific cognitive trajectories
  • Explore relationship of CDR and neuropsychological tests
  • Characterize heterogeneity of cognitive trajectories
Example from biomarker study

• Select plasma samples from subjects with pre-specified trajectories:
  1. Normal cognition over time
  2. Pre-clinical MCI: normal subjects that decline to MCI
  3. MCI stable: amnestic MCI that remain stable
  4. MCI decliners: amnestic MCI that decline to dementia
  5. Dementia with AD pathology

• Samples were *manually* annotated with these 5 categories
  • ~30 subjects were selected for each category

• Our goal
  • Visualize the trajectories to help identify potential inconsistencies or outliers
  • Develop classification algorithms to automatically label subjects
Visualization of longitudinal cognitive within each manually assigned category

<table>
<thead>
<tr>
<th>Normal</th>
<th>Pre-clinical MCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>9990 N : N : N : N</td>
<td>N : normal</td>
</tr>
<tr>
<td>10040 N : N : N : N : N : N</td>
<td>SCC: subjective cognitive concern,</td>
</tr>
<tr>
<td>10158 N : N : N : N</td>
<td>MCI_AD: amnestic MCI,</td>
</tr>
</tbody>
</table>

Example: Subject with 3 visits with N and 3 MCI_AD Dx
Visualization of longitudinal cognitive within each manually assigned category

MCI Stable

MCI Decliners

Visualizing entire longitudinal history can assist in labeling
Visualization of CDR Sum of Boxes for pre-defined categories

Cognitive Trajectory
identifying outliers. Assist in model design
Further explore pre-MCI AD

Neuropsych Test
Z-scores using NACC norms with age, sex, education adjustments were computed.

Executive: Trails B and digits backward
Language: Animals & Vegetables
Memory: Craft immediate and delayed paraphrase
How do slopes of CDR SOB trajectory relate to slopes of Neuro-Psych?
How do slopes of CDR SOB trajectory relate to slopes of Neuro-Psych?
How do slopes of CDR SOB trajectory relate to slopes of Neuro-Psych in pre-MCI?

- We investigated the relationship of CDR Slopes with Neuro-Psych test slopes for pre-MCI.
- Slopes were divided into high, med or low values.
- Concordance of CDR slopes with language and memory slopes: Subjects with high CDR slopes also have high memory, high language slopes.
- CDR slopes not concordant with MOCA or executive slopes.
How do slopes of CDR SOB trajectory relate to slopes of Neuro-Psych in MCI?

- We investigated the relationship of CDR Slopes with Neuro-Psych test slopes for all MCI.
- Slopes were divided into high, med or low values.
- Higher concordance of CDR with MOCA, memory, executive or language slopes.
Do cognitive-categories have distinct patterns of CDR Sum of Boxes trajectories?

- Mixed-effect model with subject specific intercepts and slopes were fitted to the data.
- Different cognitive trajectory have distinct slopes and intercepts.

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Intercept</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCI_D</td>
<td>29</td>
<td>0.51</td>
<td>0.16</td>
</tr>
<tr>
<td>MCI_S</td>
<td>31</td>
<td>0.49</td>
<td>0.01</td>
</tr>
<tr>
<td>N</td>
<td>34</td>
<td>0.02</td>
<td>0.001</td>
</tr>
<tr>
<td>Pre-MCI</td>
<td>10</td>
<td>0.13</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Cognitive Trajectory: MCI_D, MCI_S, N, Pre-MCI.
Can we use machine learning to predict categories?

• Decision Tree
  • Technique in supervised learning algorithm for classification
  • We can generate a tree for small number of variables (see tree on left) but need machine-learning for complex data
  • Several tree-based methods: bagging, boosting and random forest
  • We used boosting method as it is resistant to overfitting
  • Boosting builds ensemble of trees from bootstrapped datasets and improves model by fitting residuals
Performance of machine learning tool

- We used CDR slope and intercept as input variables
- Used 1/3 dataset as hold-out for testing
- Achieved > 80% accuracy against manual labels
- Adding additional neuro-psych test variables did not improve model performance

- Classification Error examples:
  - Pre-MCI AD -> N
  - Pre-MCI AD -> MCI_Decline_AD
  - MCI_Decline_AD -> Dem_AD
    - Example
  
  9400 AD : MCI_O : MCI_AD : AD : MCI_AD : MCI_AD : AD
Study heterogeneity of amnestic MCI with data-driven methods

• Next we explored whether data-driven-methods can identify clusters within the amnestic MCI group

• Used a finite mixture models for clustering trajectories

• CV errors are used to set # of latent classes in group-based trajectory analysis
Classify cognitive trajectory of all MADRC subjects

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>341</td>
</tr>
<tr>
<td>Pre-MCI_AD</td>
<td>114</td>
</tr>
<tr>
<td>MCI_Stable_AD</td>
<td>242</td>
</tr>
<tr>
<td>MCI_Decline_AD</td>
<td>200</td>
</tr>
<tr>
<td>Dem_AD</td>
<td>160</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1057</strong></td>
</tr>
</tbody>
</table>
Interactive web-based tool
Conclusions & Future Directions

• Developed machine-learning model to categorize cognitive trajectory
• Built visualization tools to illustrate trajectories
• In future too will be deployed to center researchers
  • Allow researchers to query databases for cognitive trajectories of interest
  • Add demographics, APOE genotype, longitudinal blood-draw and imaging information to visualization
Multi-disciplinary Approach

Challenge: Assess cognitive Dx trajectory using tabular format in spreadsheets

Visualization provides a manageable representation of the same data leading to more efficient labeling

Computational model further increases scalability

Model is further refined by input from clinical experts

Multidisciplinary framework can be applied to other problems
Acknowledgements

- Bradley T Hyman, MD, PhD
- Steven E Arnold, MD, PhD
- Derek Oakley, MD, PhD

- Colin Magdamo
- Emily Merrill