Comparison of methods for algorithmic classification of dementia status in the Health and Retirement Study

Kan Z. Gianattasio1, Qiong Wu2, M. Maria Glymour3, Melinda C. Power1

1Department of Epidemiology and Biostatistics, Milken Institute School of Public Health, George Washington University; 2Institute of Social Science Survey, Peking University; 3Department of Epidemiology and Biostatistics, University of California, San Francisco

FUNDING ACKNOWLEDGMENT: This work was funded by R03 AG055485

INTRODUCTION AND OBJECTIVES

- Dementia ascertainment is time-consuming and costly, thus it is difficult to describe and monitor trends and disparities in the prevalence and incidence of cognitive impairment.
- Researchers have independently developed algorithms to use existing data from the Health and Retirement Study (HRS) to algorithmically classify dementia status in cohort participants, but reporting of performance metrics is inconsistent.
- The objective of this study is to conduct a head-to-head comparison of performance of 5 existing algorithms for algorithmic classification of dementia in the HRS.

DATA AND METHODS


RESULTS

Sample Descriptions: There was a higher portion of dementia cases (34% vs. 15%), and proxy-respondents (22% vs. 6%) in the training data compared to the validation data. While training data participants also had more physical functioning limitations, the two groups were similar in sociodemographics and cognitive functioning. Algorithm Descriptions: Separate algorithms were used for self-respondents versus proxy-respondents by all authors except Wu et al., who used the missing-indicator method to combine self- and proxy-respondents into a single algorithm and setting non-applicable items to 0 (Table 1).

Table 1: Details of model choice and variables included by algorithm

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Score cutoff</td>
<td>Sens (5)</td>
<td>Spec (7)</td>
<td>Sens (11)</td>
<td>Sens (11)</td>
<td>Sens (11)</td>
</tr>
<tr>
<td>Demographics</td>
<td>Age</td>
<td>Gender</td>
<td>Race</td>
<td>Cognition (self-response)</td>
<td>Word recall</td>
</tr>
<tr>
<td>Cognition (proxy)</td>
<td>Proxy-rated memory Interviewer assess.</td>
<td>16-item IQCODE</td>
<td>7-item Jorm symps</td>
<td>Physical Functioning</td>
<td>ADLS’s</td>
</tr>
</tbody>
</table>

Table 2: Overall Performance metrics for each data sample (0.5 cut-point)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Training (N=760)</th>
<th>Validation (N=515)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sens</td>
<td>Spec</td>
<td>Acc</td>
</tr>
<tr>
<td>Herzog-Wallace</td>
<td>89.9</td>
<td>79.1</td>
</tr>
<tr>
<td>Langa-Kabeto-Weir</td>
<td>75.2</td>
<td>83.3</td>
</tr>
<tr>
<td>Wu</td>
<td>62.0</td>
<td>82.2</td>
</tr>
<tr>
<td>Hurd</td>
<td>76.7</td>
<td>91.8</td>
</tr>
</tbody>
</table>

Figure 1: ROC curves for each regression-based algorithm in the training and validation data

Figure 2: Performance metrics for each algorithm, by race/ethnicity, by sample

CONCLUSIONS AND DISCUSSION

- Higher sensitivity in the training and alternate validation data suggest that existing algorithms are better at predicting prevalent than incident dementia.
- The usefulness of each algorithm will be determined by the purpose:
  - At cut-point = 0.5, Crimmins provides highest sensitivity and Herzog-Wallace provides highest specificity, while Hurd offers highest overall accuracy.
  - Hurd also minimizes race/ethnic disparities in prevalent cases, while Wu/Crimmins minimize these disparities in incident dementia.
- The relative ease of applying these algorithms will also be a key factor to consider: regression-based algorithms are much more difficult and time-consuming to implement.

LIMITATIONS:

- We assume a time-invariant relationship between predictors and dementia.
- Validation and training data drawn from same study limits external validity.
- Validation data includes only incident cases, which are not ideal for evaluating algorithms developed with prevalent cases.
- Small N’s limit conclusiveness of sub-group differences.

FUTURE DIRECTIONS:

- Further testing of existing algorithms using external data sources, separately for prevalent and incident dementia.
- Developing improved algorithms for classifying dementia using variables commonly collected in large population surveys, with a particular focus on achieving uniform performance across subgroups.