RISK PREDICTION OF A MULTIPLE SCLEROSIS DIAGNOSIS

Joyce C. Ho, Joydeep Ghosh, K.P. Unnikrishnan
Multiple Sclerosis Introduction

- Chronic, progressive, and incurable immune disease
- Nerve damage causes signal disruption in brain and spinal cord
- 2.5 million people worldwide suffer from MS
- Early diagnosis can slow progression and prolong a “normal” lifestyle
Related Work

• Epidemiology studies

• Predictive models
Objective

Can EMR-based risk prediction model help medical professionals diagnosis MS?
NorthShore EMR Data

• De-identified patient data from NorthShore Enterprise Data Warehouse between Jan 2006 – July 2012
  - Demographics
  - Family medical history
  - Medical diagnosis
  - Vaccination history
  - Blood test results

Derived from ICD-9
NorthShore MS Data

- Case patients identified using MS ICD-9 codes during neurology visit
- Age and gender matched case-control
- 3,685 patients – 20% prevalence
- Encounters after diagnosis date excluded
  - Control: random encounter date
  - Case: initial MS diagnosis
Encounters Prior to Diagnosis

limited medical history for MS patients
Feature Construction

- Complete demographic info required
- Family history, medical diagnoses, vaccination history are binarized
- Blood test results encoded as observed-abnormal, observed-normal, unobserved
- Total 56 features
## EMR Features

<table>
<thead>
<tr>
<th>Set</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Demographic + Family medical history</td>
</tr>
<tr>
<td>2</td>
<td>Feature Set 1 + Autoimmune diagnoses</td>
</tr>
<tr>
<td>3</td>
<td>Feature Set 2 + Microbial diagnoses</td>
</tr>
<tr>
<td>4</td>
<td>Feature Set 3 + Mental illness diagnoses</td>
</tr>
<tr>
<td>5</td>
<td>Feature Set 4 + Cancer diagnoses</td>
</tr>
<tr>
<td>6</td>
<td>Feature Set 5 + Vaccination history</td>
</tr>
<tr>
<td>7</td>
<td>Feature Set 6 + Reproductive related codes</td>
</tr>
<tr>
<td>8</td>
<td>Feature Set 7 + MRI scans + obesity</td>
</tr>
<tr>
<td>9</td>
<td>Feature Set 8 + Blood tests results</td>
</tr>
</tbody>
</table>
Experimental Setup

• 10-fold cross validation
• Feature selection with forward and backward stepwise regression using Akaike information criterion
• Multivariate logistic regression for classification
RESULTS

Comparison of EMR Features
Feature Set Performance

- + autoimmune diagnosis
- + vaccination history
- + blood tests

AUC vs Feature Set

Data at first visit

+ vaccination history
Feature Set Selection

features are more informative

< 15% of features selected
Case Transition Probabilities

- Slight improvement
  - Autoimmune diagnosis
  - Vaccination history
  - Blood tests

Noticeable probability increase
Control Transition Probabilities

- + autoimmune diagnosis
- + vaccination history
- + blood tests

Minimal changes in prediction
Gradual decrease in risk
Improved prediction of control patients
RESULTS

Using All Features with Forward Selection Model
Predicted Risk Probabilities

![Predicted Risk Probabilities Graph](image-url)
ROC Performance

Significant improvement using all features

Data available at first visit barely beats random chance
Sensitivity, Specificity, PPV Tradeoffs

Sensitivity = Specificity = 0.528
PPV = 0.314

Feature Set 1

Sensitivity = Specificity = 0.647
PPV = 0.314
## Logistic Regression Coefficients

<table>
<thead>
<tr>
<th>Feature</th>
<th>Beta</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSF oligoclonal bands (present)</td>
<td>16.255±0.545</td>
<td>0.000</td>
</tr>
<tr>
<td>Mental illness (FH)</td>
<td>6.298±3.101</td>
<td>0.033</td>
</tr>
<tr>
<td>Epstein-Barr Virus</td>
<td>3.974±3.924</td>
<td>0.093</td>
</tr>
<tr>
<td>Abnormal brain MRI</td>
<td>2.877±0.313</td>
<td>0.000</td>
</tr>
<tr>
<td>B12 blood test (unobserved)</td>
<td>2.527±0.149</td>
<td>0.047</td>
</tr>
<tr>
<td>HPV Vaccine</td>
<td>-15.728±2.188</td>
<td>0.000</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>-15.763±0.369</td>
<td>0.235</td>
</tr>
<tr>
<td>Estrogen replacement</td>
<td>-15.823±0.209</td>
<td>0.037</td>
</tr>
<tr>
<td>Inflammatory bowel disease (FH)</td>
<td>-17.236±0.420</td>
<td>0.885</td>
</tr>
<tr>
<td>HIB vaccine</td>
<td>-18.156±3.101</td>
<td>1.000</td>
</tr>
</tbody>
</table>
CONCLUSIONS
Discussion

- Usage of ICD-9 codes for patient’s history
  - Billing data has poor sensitivity and specificity
- Limited patient history for case patients
- Clinical notes review may produce additional, more informative features
Summary

- Model to identify patients at high-risk of developing MS
- Sparse set of features obtains reasonable predictive performance
- Demonstration of leveraging EMRs to aid medical professionals with early disease diagnosis