KEEPING AN EYE ON HEALTHCARE COSTS

The D2Hawkeye Story

Sparrowhawk photo courtesy Mike Atkins on Wikimedia Commons. License: CC BY.
D2Hawkeye

- Founded by Chris Kryder, MD, MBA in 2001

- Combine expert knowledge and databases with analytics to improve quality and cost management in healthcare

- Located in Massachusetts USA, grew very fast and was sold to Verisk Analytics in 2009
D2Hawkeye

Data Sources
- Diagnoses
- Procedures
- Drugs
- ...

Aggregate
Clean
Normalize

Secure
Database

Reports

Predictive
Models
Healthcare Case Management

- D2Hawkeye tries to improve healthcare case management
  - Identify high-risk patients
  - Work with patients to manage treatment and associated costs
  - Arrange specialist care

- Medical costs often relate to severity of health problems, and are an issue for both patient and provider

- Goal: improve the quality of cost predictions
Impact

- Many different types of clients
  - Third party administrators of medical claims
  - Case management companies
  - Benefit consultants
  - Health plans

- **Millions of people** analyzed monthly through analytic platform in 2009

- **Thousands of employers** processed monthly
Pre-Analytics Approach

- Human judgment – MDs manually analyzed patient histories and developed
- Limited data sets
- Costly and inefficient
- Can we use analytics instead?
Data Sources

• Healthcare industry is data-rich, but data may be hard to access
  • Unstructured – doctor’s notes
  • Unavailable – hard to get due to differences in technology
  • Inaccessible – strong privacy laws around healthcare data sharing

• What is available?
Data Sources

- Claims data
  - Requests for reimbursement submitted to insurance companies or state-provided insurance from doctors, hospitals and pharmacies.

- Eligibility information

- Demographic information
## Claims Data

<table>
<thead>
<tr>
<th>ClaimType</th>
<th>ProviderName</th>
<th>DiagCode</th>
<th>DiagDesc</th>
<th>Source</th>
<th>SourceDiagDesc</th>
<th>ProcNDC Code</th>
<th>ProcNDCDesc</th>
<th>ServiceDate</th>
<th>PaidAmount</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEN</td>
<td>SOUTHEASTERN MINNESOTA ORAL &amp; MAX ASSOCIATED ORAL &amp; MAXILLOFACIAL SURGEONS PA CENTRAL FLORIDA ORAL SURGERY</td>
<td>DD0238</td>
<td>Dental Diseases</td>
<td>DD0238</td>
<td>Unspecified Anomaly of Tooth Position</td>
<td>DD007</td>
<td>Anesthesia - General</td>
<td>4/22/2005</td>
<td>$ -</td>
</tr>
<tr>
<td>DEN</td>
<td>CENTRAL FLORIDA ORAL SURGERY</td>
<td>DD0238</td>
<td>Dental Diseases</td>
<td>DD0238</td>
<td>Disturbances in ToOther Eruption</td>
<td>DD025</td>
<td>Dental</td>
<td>7/8/2005</td>
<td>$ 272.68</td>
</tr>
<tr>
<td>DEN</td>
<td>CENTRAL FLORIDA ORAL SURGERY</td>
<td>DD0238</td>
<td>Dental Diseases</td>
<td>DD0238</td>
<td>Disturbances in ToOther Eruption</td>
<td>DD025</td>
<td>Dental</td>
<td>11/11/2005</td>
<td>$ 568.13</td>
</tr>
<tr>
<td>Med</td>
<td>ALPHARETTA INTERNA</td>
<td>DD0004</td>
<td>ENT and Upper Resp Disorders</td>
<td>DD0004</td>
<td>Acute Maxillary Sinusitis</td>
<td>DD147</td>
<td>Office Visit - Established Patient</td>
<td>5/26/2005</td>
<td>$ 125.85</td>
</tr>
<tr>
<td>Med</td>
<td>CUMMING FAMILY MEDICINE</td>
<td>DD0170</td>
<td>Neurotic and Personality Disorders</td>
<td>DD0170</td>
<td>Neurotic Disorders</td>
<td>DD149</td>
<td>Office Visit - New Patient</td>
<td>6/20/2005</td>
<td>$ -</td>
</tr>
<tr>
<td>Med</td>
<td>ATLANTA WOMENS HEALTH GROUP-582483738.20</td>
<td>DD0102</td>
<td>Screening</td>
<td>V776</td>
<td>Special Screening for Cystic Fibrosis</td>
<td>DD077</td>
<td>Lab - Blood Tests</td>
<td>7/29/2005</td>
<td>$ 1.52</td>
</tr>
</tbody>
</table>
Claims Data

• Rich, structured data source
• Very high dimension
• Doesn’t capture all aspects of a person’s treatment or health – many things must be inferred
• Unlike electronic medical records, we do not know the results of a test, only that a test was administered
D2Hawkeye’s Claims Data

- Available: claims data for 2.4 million people over a span of 3 years

- Include only people with data for at least 10 months in both periods – 400,000 people
Variables

13,000 diagnoses  →  217 diagnosis groups

22,000 procedures  →  213 procedure groups

45,000 prescription drugs  →  189 therapeutic groups
Variables – Cost Profiles

- Patient 1
- Patient 2

15.071x – Keeping an Eye on Healthcare Costs: The D2Hawkeye Story
Additional Variables

• Chronic condition cost indicators

• 269 medically defined risk rules
  • Interactions between illnesses
  • Interactions between diagnosis and age
  • Noncompliance to treatment
  • Illness severity

• Gender and age
Cost Variables

- Rather than using cost directly, consider every $1000:
  - $8000-$19000: 5.5%
  - $19000-$55000: 2.1%
  - $3000-$8000: 13.8%
  - $55000+:
    - <$3000: 78.0%
    - $3000-$8000: 13.8%
Medical Interpretation of Buckets

Bucket 1: Low
- Candidate for Wellness Programs

Bucket 2: Emerging
- Candidate for Disease Management Programs

Bucket 3: Moderate

Bucket 4: High
- Candidate for Case Management

Bucket 5: Very High

15.071x – Keeping an Eye on Healthcare Costs: The D2Hawkeye Story
Error Measures

• Typically we use $R^2$ or accuracy, but others can be used

• In case of D2Hawkeye, failing to classify a high-cost patient correctly is worse than failing to classify a low-cost patient correctly

• Use a “penalty error” to capture this asymmetry
Penalty Error

- Key idea: use asymmetric penalties
- Define a “penalty matrix” as the cost of being wrong

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 2 4 6 8</td>
</tr>
<tr>
<td>2</td>
<td>1 0 2 4 6</td>
</tr>
<tr>
<td>3</td>
<td>2 1 0 2 4</td>
</tr>
<tr>
<td>4</td>
<td>3 2 1 0 2</td>
</tr>
<tr>
<td>5</td>
<td>4 3 2 1 0</td>
</tr>
</tbody>
</table>
Baseline

• Baseline is to simply predict that the cost in the next “period” will be the cost in the current period

• Accuracy of 75%

• Penalty Error of 0.56
Multi-class Classification

- We are predicting a bucket number
- Example

Diagram:
- C.A.D.
- No C.A.D.
- Diabetes
- No diabetes
- 5
- 3
- 1
Most Important Factors

• First splits are related to cost

```
Paid ≤ $4000

Paid > $4000

Paid ≤ $40000

Paid > $40000
```

15.071x – Keeping an Eye on Healthcare Costs: The D2Hawkeye Story
Secondary Factors

- Risk factors
- Chronic Illness
- “Q146”
  - Asthma + depression
- “Q1”
  - Risk factor indicating hylan injection
  - Possible knee replacement or arthroscopy
Example Groups for Bucket 5

• Under 35 years old, between $3300 and $3900 in claims, C.A.D., but no office visits in last year

• Claims between $3900 and $43000 with at least $8000 paid in last 12 months, $4300 in pharmacy claims, acute cost profile and cancer diagnosis

• More than $58000 in claims, at least $55000 paid in last 12 months, and not an acute profile
## Results

<table>
<thead>
<tr>
<th>Bucket</th>
<th>Trees</th>
<th>Baseline</th>
<th>Trees</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>80%</td>
<td>75%</td>
<td>0.52</td>
<td>0.56</td>
</tr>
<tr>
<td>1</td>
<td>85%</td>
<td>85%</td>
<td>0.42</td>
<td>0.44</td>
</tr>
<tr>
<td>2</td>
<td>60%</td>
<td>31%</td>
<td>0.89</td>
<td>0.96</td>
</tr>
<tr>
<td>3</td>
<td>53%</td>
<td>21%</td>
<td>1.01</td>
<td>1.37</td>
</tr>
<tr>
<td>4</td>
<td>39%</td>
<td>19%</td>
<td>1.01</td>
<td>1.72</td>
</tr>
<tr>
<td>5</td>
<td>30%</td>
<td>23%</td>
<td>1.01</td>
<td>1.88</td>
</tr>
</tbody>
</table>
Insights

• Substantial improvement over the baseline

• Doubled accuracy over baseline in some cases

• Smaller accuracy improvement on bucket 5, but much lower penalty
Analytics Provide an Edge

• Substantial improvement in D2Hawkeye’s ability to identify patients who need more attention

• Because the model was interpretable, physicians were able to improve the model by identifying new variables and refining existing variables

• Analytics gave D2Hawkeye an edge over competition using “last-century” methods