Predicting future resource use & risk of hospitalization for a general population in NHS England: Adapting US models & potential lessons for the US

Stephen Sutch
Johns Hopkins Bloomberg School of Public Health

To be presented at The Predictive Modeling Summit
Washington, DC, November 14, 2014
Introduction

• A number of models are available in the US and the UK which predict the risk of hospitalisation, from general and insured populations.

• Multiple purposes e.g. screening of patients for Case Management Programs, screening for Disease Management Programs, organisational profiling, and assessing financial risk.

• Response to health policies to reduce unnecessary hospital admissions, Pay for Performance (P4P) measures, Risk stratification tool requirements.

• A need to support populations in avoiding hospital admissions that are both expensive and a patient safety risk.
Historic Use of Models in England

- Existing predictive models in the ACG System were based on US data, rescaled on local data.
- Early work at Imperial College and UCL showed the applicability of the ACG System to NHS data.
- In 2006, Johns Hopkins University and the Kings Fund created predictive models from NHS data.
- Leeds City PCT showed existing models in ACG System could match and exceed the performance of the Combined Predictive Model (CPM).
- Currently used in NHS to create lists of individuals for clinical review, care management to prevent unnecessary hospital admissions.
Role of Clinical Commissioning Groups (CCG)

- “Planning services based on the needs of the local population”
- “Securing services that meet the needs of the local population”
- “Monitoring the quality of care provided”

- 2013 - 211 CCGs (avg 226k pop, 60% of total NHS budget)
- “All GP (PCP) practices have to be members of a CCG, and every CCG board will include at least one hospital doctor, nurse and member of the public.”

Source: http://www.patient.co.uk/
Using Predictive Modeling to Assign Persons Within the Care Management Pyramid

- **5% Level 3**
  - High risk with multiple chronic illness
  - Intensive Case and Disease Management

- **15% Level 2**
  - Moderate risk patients with single chronic illness or risk factors
  - Health Coaching and Lifestyle Management

- **80% Level 1**
  - Low risk
  - Health Education and Promotion
ACG System predictive models used to generate an outreach “list” for GPs, care management nurses / Community Matrons

<table>
<thead>
<tr>
<th>Patient Id</th>
<th>Age</th>
<th>Sex</th>
<th>Total Cost</th>
<th>Rescaled Total Cost</th>
<th>Resource Index</th>
<th>Probability High Total Cost</th>
<th>Probability IP Hospitalization</th>
<th>Unique Provider Count</th>
<th>Rx Gaps</th>
<th>Hospital Dominant Count</th>
<th>Chronic Condition Count</th>
<th>Frailty Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>41316*141931615</td>
<td>75 M</td>
<td>9,667.77</td>
<td>23.37</td>
<td>0.95</td>
<td>0.50</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3849996*8211951</td>
<td>55 M</td>
<td>158,168.03</td>
<td>23.73</td>
<td>0.95</td>
<td>0.50</td>
<td>6</td>
<td>0</td>
<td>5</td>
<td>9</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>411482*16139717</td>
<td>69 M</td>
<td>75,228.16</td>
<td>26.85</td>
<td>0.95</td>
<td>0.90</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>23</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3849979*3141950</td>
<td>57 M</td>
<td>281,738.71</td>
<td>25.79</td>
<td>0.95</td>
<td>0.74</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>10</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41211*161966117</td>
<td>45 M</td>
<td>38,582.56</td>
<td>29.98</td>
<td>0.95</td>
<td>0.57</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>18</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3849519*951945</td>
<td>61 M</td>
<td>21,723.69</td>
<td>25.70</td>
<td>0.95</td>
<td>0.86</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41364*14195814</td>
<td>48 F</td>
<td>62,005.22</td>
<td>28.04</td>
<td>0.95</td>
<td>0.99</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>18</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>521425*1449496</td>
<td>57 M</td>
<td>211,501.05</td>
<td>23.30</td>
<td>0.95</td>
<td>0.78</td>
<td>5</td>
<td>8</td>
<td>2</td>
<td>22</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>414137*16194586</td>
<td>61 F</td>
<td>28,786.14</td>
<td>22.83</td>
<td>0.95</td>
<td>0.62</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>17</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>521521*14195238</td>
<td>54 F</td>
<td>116,792.19</td>
<td>22.99</td>
<td>0.95</td>
<td>0.79</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>19</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>414447*14195181</td>
<td>62 F</td>
<td>76,029.86</td>
<td>22.36</td>
<td>0.95</td>
<td>0.43</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>23</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>521565*13197135</td>
<td>55 F</td>
<td>54,057.82</td>
<td>23.05</td>
<td>0.95</td>
<td>0.50</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>20</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>414474*14194914</td>
<td>67 M</td>
<td>40,321.38</td>
<td>23.43</td>
<td>0.95</td>
<td>0.90</td>
<td>12</td>
<td>8</td>
<td>0</td>
<td>22</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>522752*161942128</td>
<td>63 M</td>
<td>72,871.18</td>
<td>23.69</td>
<td>0.95</td>
<td>0.64</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>21</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>418153*141920127</td>
<td>85 F</td>
<td>19,894.41</td>
<td>25.20</td>
<td>0.95</td>
<td>0.85</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>25</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>522163*141951127</td>
<td>55 M</td>
<td>121,288.81</td>
<td>30.24</td>
<td>0.95</td>
<td>0.93</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>34</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>416334*16193110</td>
<td>66 M</td>
<td>66,142.02</td>
<td>31.58</td>
<td>0.95</td>
<td>0.95</td>
<td>9</td>
<td>0</td>
<td>5</td>
<td>29</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>526154*16194635</td>
<td>63 F</td>
<td>86,538.62</td>
<td>29.92</td>
<td>0.95</td>
<td>0.45</td>
<td>12</td>
<td>11</td>
<td>4</td>
<td>26</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>415623*14195516</td>
<td>49 F</td>
<td>61,451.14</td>
<td>25.72</td>
<td>0.95</td>
<td>0.97</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>26</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>452121*141942125</td>
<td>54 M</td>
<td>125,962.55</td>
<td>42.58</td>
<td>0.95</td>
<td>0.90</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>35</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>417446*171861027</td>
<td>19 M</td>
<td>220,751.36</td>
<td>38.46</td>
<td>0.95</td>
<td>0.94</td>
<td>8</td>
<td>0</td>
<td>5</td>
<td>30</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>564541*141938301</td>
<td>68 F</td>
<td>96,138.29</td>
<td>22.96</td>
<td>0.95</td>
<td>0.92</td>
<td>9</td>
<td>0</td>
<td>3</td>
<td>19</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>442562*141931299</td>
<td>70 F</td>
<td>67,588.26</td>
<td>37.32</td>
<td>0.95</td>
<td>0.70</td>
<td>7</td>
<td>2</td>
<td>4</td>
<td>21</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>456729*141950126</td>
<td>55 F</td>
<td>11,361.46</td>
<td>23.86</td>
<td>0.95</td>
<td>0.36</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>444441*161927103</td>
<td>78 F</td>
<td>67,767.97</td>
<td>29.27</td>
<td>0.95</td>
<td>0.71</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>12</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Comprehensive Patient Clinical Profile

### Summary

**Comprehensive Patient Clinical Profile Report - Patient Id: 7442522*16195151**

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>PCP Id</th>
<th>Product</th>
<th>Resource Utilization Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>M</td>
<td>5212*11</td>
<td>HMO</td>
<td>5</td>
</tr>
</tbody>
</table>

### Model

- **DxRxPM - total cost - incident dx -> total cost**
- **DxRxPM - rx cost - incident dx -> rx cost**
- **Prior Costs**
  - Total Cost: $26,951
  - Rx Cost: $2,700

### Predictive Values

- **Probability High Total Cost**: 0.65
- **Predicted Total Cost Range**: $30,000-$40,000
- **Probability High Rx Cost**: 0.44
- **Predicted Rx Cost Range**: $2,000-$3,000
- **High Risk Unexpected Pharmacy**: N

### Coordination of Care

- **Chronic Condition Count**: 12
- **Unique Providers Seen**: 2
- **Specialty Types Seen**: 2
- **No Generalist Seen**: Y
- **% Visits Provided By Majority Source of Care**: 67
- **Frailty Flag**: N

### Utilization

- **Outpatient Visits**: 55
- **ER Visits**: 3
- **Inpatient Admissions**: 0
- **Major Procedure Performed**: Y
- **Diagnosis Service**: N
- **Nursing Service**: N

### Likelihood of Hospitalization

- **Hospital Dominant Count**: 3
- **Probability Hospital Admission (6 mos)**: 0.04
- **Probability Hospital Admission (12 mos)**: 0.07
- **Probability ICU/CCU Admission**: 0.01
- **Probability Injury-related Admission**: 0.02
- **Probability Long-term Admission (12 + days)**: 0.01

### Condition Profile with Pharmacy Adherence

**Condition** | **Present?** | **CSA** | **MPR** | **# Refill Gaps** | **Untreated**
--- | --- | --- | --- | --- | ---
Age-Related Macular Degeneration | NP | | | | |
Bipolar Disorder | | | | | |
Congestive Heart Failure | NP | | | | |
Depression | ICD | | | | |
Diabetes | ICD | | | | |
Glaucoma | ICD | | | | |
Human Immunodeficiency Virus | NP | | | | |
Disorders of Lipid Metabolism | Rx | | | | |
Hypertension | ICD | | | | |
Hyperthyroidism | ICD | | | | |
Immunosuppression/Transplant | NP | | | | |
Ischemic Heart Disease | NP | | | | |
Osteoporosis | | CSA | | | |
Parkinson’s Disease | NP | | | | |
Persistent Asthma | Rx | | | | |
Rheumatoid Arthritis | NP | | | | |
Schizophrenia | ICD | | | | |
Seizure Disorders | NP | | | | |
COPD | ICD | | | | |
Chronic Renal Failure | NP | | | | |
Low Back Pain | ICD | | | | |

### High Impact Conditions

**EDCs**

- GAS02: Inflammatory bowel disease
  - Rx/MGs: RES020: Respiratory / Chronic Medical

### Moderate Impact Conditions

**EDCs**

- CAR14: Hypertension, w/o major complications
  - Rx/MGs: CAR040: Cardiovascular / Disorders of Lipid

### Low Impact Conditions

**EDCs**

- ADM02: Surgical illness
  - Rx/MGs: GSI030: General Signs and Symptoms / Pain and
- ADM05: Administrative concerns and non-specific
  - Rx/MGs: INF020: Infections / Acute Minor
- ADM08: Preventive care
  - Rx/MGs: ZZZ000: Other and Non-Specific Medications
- ALL03: Allergic rhinitis
Example Clinical Process

- Identify at risk patients – ACG risk profiling tool
- Core medical team review
  - Identify problems, Action list, Suitability for further interventions
- Personalized care plan
  - Discussion and delivery of care plan, Coded and scanned to records
- Follow-up
  - Clinical review (named clinician), Date of review, Response to interventions

Source: Cricket Green Medical Practice Model
The South Central Region of the NHS

- 9 primary care trusts (PCTs)
- 510 GP practices clustered into 20 CCGs
- 4 million population

PCTs currently responsible for commissioning of services

ACGs in use in approximately half of GP practices
A complex end-to-end infrastructure that took over 9 months to put in place but:

- It addresses all of the issues/concerns/requirements of our stakeholder group particularly around the issue of transferring, storing and sharing data, particularly primary care data
- Primary care data extraction – a complex and resource intensive process - is undertaken by a specialist company rather than PCT staff
- End users have access to a user-friendly graphical interface on their desktop
- It only takes 4-6 weeks from a GP practice opting in and having access to ACG information
Method

• Aim: apply the ACG System variables as independent variables in year 1, to predict patient outcomes in year 2

• Two main dependent (outcome) variables,
  – total cost in year 2 (Linear Regression)
  – hospitalization in year 2 (Logistic Regression)

• Objectives
  – create predictive models from English NHS data
  – validate those models (split half validation)
  – compare with the existing US-based models
  – recommend a model for application England.
Risk Factors in the Johns Hopkins Predictive Model

- Overall Disease Burden (Dx → ACG)
- Selected Medical Conditions (Dx → Expanded Dx Clusters)
- Special Population Markers (Dx → HOSDOM, Frailty)
- Medications (Rx → Rx-MG)
- Selected Resource Use Measures ($)
- Age
- Gender
## Results (1)

- Data: 663,797 individuals in year 1
- extracted from primary care practices which had completed and approved a consent process.
- Secondary care data was added from hospital data for cases where patients had also received hospital services.
- Linear regression to predict future (year 2) total patient expenditure, R-Square 27.5% untrimmed
- R2 8.8% age/gender, 22.4% US based models
- With prior cost and utilisation variables added the model’s performance increased to 30.9%
R Squared Results

<table>
<thead>
<tr>
<th>Age / Gender</th>
<th>ACG w/o Prior</th>
<th>ACG w Prior</th>
<th>ACG US All Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>.0910</td>
<td>.2792</td>
<td>.3010</td>
</tr>
<tr>
<td>Validation</td>
<td>.0902</td>
<td>.2745</td>
<td>.2943</td>
</tr>
<tr>
<td></td>
<td>12 Month Admission</td>
<td>6 Month Admission</td>
<td>&gt;12 day LoS</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>Train</strong></td>
<td>0.795</td>
<td>0.814</td>
<td>0.915</td>
</tr>
<tr>
<td><strong>Validation</strong></td>
<td>0.795</td>
<td>0.815</td>
<td>0.904</td>
</tr>
</tbody>
</table>
Results (2)

- logistic model to predict unplanned hospitalization
  - C-Statistic 0.78
  - Directly related to measure used in P4P program for PCPs (NHS QoF)
    - Reduction in avoidable hospital admissions
  - “Emergency Admissions” (3.74%)
## Risk of Unplanned Admission (3.74%)
Sensitivity / PPV, NHS England 2013

<table>
<thead>
<tr>
<th>Split</th>
<th>Cut Pt</th>
<th>Sens</th>
<th>Spec</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>.0198</td>
<td>84.6%</td>
<td>51.3%</td>
<td>6.3%</td>
<td>98.90%</td>
</tr>
<tr>
<td>90%</td>
<td>.0790</td>
<td>43.66%</td>
<td>91.28%</td>
<td>15.99%</td>
<td>97.71%</td>
</tr>
<tr>
<td>95%</td>
<td>.1227</td>
<td>29.91%</td>
<td>95.95%</td>
<td>21.90%</td>
<td>97.30%</td>
</tr>
<tr>
<td>98%</td>
<td>.2048</td>
<td>16.56%</td>
<td>98.55%</td>
<td>30.31%</td>
<td>96.88%</td>
</tr>
<tr>
<td>99%</td>
<td>.2874</td>
<td>10.30%</td>
<td>99.35%</td>
<td>37.71%</td>
<td>96.68%</td>
</tr>
<tr>
<td>99.5%</td>
<td>.3817</td>
<td>6.00%</td>
<td>99.71%</td>
<td>43.89%</td>
<td>96.54%</td>
</tr>
</tbody>
</table>
Discussion (1)

- The results show a statistically significant improvement over the existing models available in the ACG System implemented in the UK NHS, consistent with similar projects carried out in Sweden and Spain.
- The original US models still provided good sufficient estimates that have been proven to be robust in a number of countries over several decades.
Conclusion (1)

- Casemix classifications reduce data complexity and provide robust measures of multimorbidity. The models work well in explaining the top 1% and 5% of data, but also perform well in discriminating risk “lower in the population pyramid” to identify potential emerging risk.

- Current emphasis on identifying the highest risk individuals, there is an increased interest in recognising earlier and emerging risk, where more preventative methods can be informed such as chronic disease self-management programs.
Conclusion (2)

- A standard set of independent variables were used in the models. Additional variables could be used in future models such as BMI, Smoking Status, and social care data.
- Alternative models can produce higher results by using current utilisation and costs measures, however these models would increase bias to individuals already accessing healthcare services to the detriment of those with low current access.
- Including prior utilisation and prior cost measures as independent variables also creates perverse incentives to increase resource use.”
Discussion (1)

• **Intermediate Classification**
  - Form a set of independent variables from 1000s of input variables
  - Dependent Variable, move from Any admission to unplanned/emergency/preventable

• **Additional Variables, Data**
  - Additional variables could be used in future models such as BMI, Smoking Status, and social care data.

• **Alternative models needed**
  - Historic utilization can produce higher results but bias to individuals already accessing healthcare
  - creates perverse incentives to increase resources
  - Dependent variable, Unplanned admissions
Discussion (2)

• Creating alternative Views
  – Concurrent v Prospective (Performance measurement v Planning)
  – Individuals, Populations

• Longitudinal data, Changing Risk
  – Increasing, decreasing, see-sawing
  – Real-time alerts

• EHR and Social Data
  – Data linkage, assessments, labs
  – Patient data - Health Status, Behaviour, Self-Assessment (e.g. SF12/36, EQ5D, PAM, HRA, PHQ9)
  – Selection Bias (Non-response, Exclusion bias)
Opportunities for Learning more....

• Web Site:
  – www.acg.jhsph.edu

• Contact:
  – Steve Sutch, Dir. Product Management, ACG International
    ssutch1@jhu.edu
Results - Hospitalisation

- logistic model to predict future hospitalisation
  - C-Statistic 0.80
  - age/gender model 0.67
  - current US model 0.75
  - For purposes of generating lists of high risk individuals applying a cut-point such that 1% of the population are designated as “positive”, the model showed a positive predictive value of 65.46%