Surveillance of Hospital Acquired Infections Using EMR Data

Richard Davies, MD, PhD

PHO Grand Rounds

January 22, 2013
• No conflicts of interest to declare related to this work.
Acknowledgements

• Funding
  – Center for Security Science, Defense Research Canada

• Lead Federal Department
  – National Research Council of Canada Institute for Information Technology

• Lead Industry Partner
  – AMITA Corporation, Ottawa

• Collaborators
  – Health Care
  – Computer Science
  – Statistics and Modeling
  – Epidemiology and Public Health
  – Mapping
  – Privacy
Collaborators

• Health Care
  – Ottawa Heart Institute
  – All Ottawa Area Hospitals
  – All Grey Bruce Area Hospitals
  – Kingston General Hospital
  – Infectious Disease Divisions
  – Emergency Medicine Departments

• Computer Science
  – NRC – IIT
  – AMITA Corp (Ottawa)
  – TOH IT
  – TOH Data Warehouse
  – Altarum Corp (Michigan)

• Statistics and Modeling
  – STATA Corp. (Texas)
  – Carnegie Mellon University Dept. of Computer Science
  – DRDC Valcartier

• Mapping
  – City of Ottawa IT Department
  – DM Solutions (Ottawa)

• Epidemiology and Public Health
  – Ottawa Public Health
  – Public Health Ontario
  – Grey Bruce Public Health
  – Toronto Public Health
  – York Public Health
  – Michigan Dept. of Community Health
  – KFLA Public Health/Queens University Public Health Informatics
  – Public Health Agency of Canada/CNPNI
  – Health Canada – Illicit Substances

• Privacy
  – CHEO Research Institute
  – TOH and OHI Research Ethics Boards
  – TOH, OHI and OPH Privacy Officers
Data Fusion Project

• Develop re-usable framework for surveillance
• Process
  – Deal with non technical factors
  – Data ownership access and sharing, privacy, stakeholder alignment, engagement of Subject Matter Experts
• Technical/software
  – Data Management
  – Data Science
  – Statistics and Visualization
• Demonstrate using two retrospective test scenarios
  – Hospital Acquired Infections
  – Harm Related to Illicit Drugs
Process Framework

• Stakeholder identification and alignment
  – Data owners
  – Researchers who may see you as competing with them
  – Hospital administration

• SME engagement (public health, infectious diseases)
  – Problem definition
  – Syndrome definitions
  – Useful data display and output

• Data sharing agreements

• Privacy, REB approvals
Data Fusion Technical Framework

- EMR Data Streams
- Data Management
- Data Science
- Subject Matter Experts
- Statistics and Visualization
- Useful Output

Diagram showing the flow of information from EMR Data Streams to Data Management, then to Data Science, followed by Subject Matter Experts and finally to Useful Output, with intermediate connections to Statistics and Visualization.
Data Fusion Technical Framework: Data Management

- MIRTH Connect HL-7 Compliant Enterprise Database Management System (EDMS)
- Maintain multidimensional data from different sources in linkable format
- Maintain data integrity and security
- Controlled access: Allow research while protecting privacy
- Output customized data files optimized for analysis
Data Science

- Science of dealing with complex multidimensional data
- Term coined in 1960s
- Proposed as independent area in 2001
- Eight “parent” disciplines (opposite)
- Translating data in its existing format into useful information.
- Research areas include
  - Cloud computing
  - Information retrieval
  - Databases and information integration
  - Computer learning, NLP and information extraction
  - Knowledge discovery in social/information networks
Data Fusion Technical Framework: Data Science

• Classification and Classification of EMR data
  – Natural Language Programming (NRC and OHI)
• “Syndrome” Definition
  – = Quantitative variable derived from data
  – Relevant to problem
  – Simple syndrome: derived from one data element
  – Complex syndrome: derived by combining 2 or more elements
• Extensive SME input
Data Fusion Technical Framework: Statistics and Visualization

• Done on customized data files generated by EDMS
  – Usually spreadsheet format “flat” files.

• Traditional Statistics
  – STATA

• Mapping
  – DM Solutions
  – City of Ottawa

• Specialized Statistics
  – DRDC Val Cartier
  – Carnegie Mellon
Approach

• Best available technological components
  – Modular
  – COTS if possible
  – Scalable
  – Generalizable

• Problem rather than data driven approach
  – Focus on important problems
  – Extensive involvement of content experts
  – Selective approach to data
  – Depth rather than breadth
  – Aim for early success
Hospital Acquired Infections
Safety indicators and targets for prevention

• Ontario Ministry of Health and Long Term Care
  – Mandatory Reporting as Quality Indicators
  – Manual Data Capture
  – Public Reporting on Web

• Infections Reported
  – Ventilator Acquired Pneumonia (VAP)
  – Central Line Infections (CLI)
  – Methicillin Resistant Staph Aureus (MRSA)
  – Clostridium Difficile (C. Diff.)
  – Vancomycin Resistant Enterococcus (VRE)
Ventilator Acquired Pneumonia (VAP)

- Leading cause of death among hospital-acquired infections
  - Adds 6-30% to mortality of ventilated patients.
- Prolongs ventilator time by 4-32 days.
- Prolongs ICU LOS by 4-7 days and hospital LOS by 10 days.
- Cost estimate (US) = $10-16,000 US per VAP.
- Estimate (Canada) savings with prevention of one VAP - 14,000$.
- Estimated 4,000 cases per year in Canada
  - 230 deaths
  - 17,000 ICU days
  - 2% of all ICU days
  - $46 million per year
- Many VAP thought to be preventable.
Central Line Infections (CLI)

- Central lines
  - Commonly used
  - 48% of all ICU patients in US
  - > 90% of ICU patients in UOHI
  - Benefit
    - Long term venous access
    - Temporary Dialysis (CRRT, Hemodialysis)
    - Necessary to administer some drugs
    - Hemodynamic monitoring (Swan Ganz, CVP)
    - Includes therapeutic devices (temporary pacemakers, Intra Aortic Balloon Pump)

- Risk
  - Portal for infection

- Catheter-Related Blood Stream Infection (= Central Line Infection, CLI)
  - Attributable mortality 4-20%, estimated 500-4000 annual deaths in US (2009)
  - Prolong hospitalization by 7 days, estimated cost $3700 - $29000 US (2009)

- Potentially preventable
  - Better insertion and maintenance techniques
  - Removal as soon as not needed
Data Fusion Project: Test Data Sets

- **University of Ottawa Heart Institute**
  - 4 years of sequential admissions

- **Kingston General Hospital**
  - 3 years of sequential admissions

- **City of Ottawa Ambulance Call Reports (ACRs)**
  - 3 years of sequential data

- **Ottawa Region ER Visits**
  - Ongoing as part of ASSET Syndromic Surveillance System
Methods

• Data inputs from TOH Data Warehouse
  – Admission Discharge Transfer
    • Admissions, bed moves, physician changes, discharge/death
  – CXR reports
    • Dictated and transcribed report on every CXR done
  – Pharmacy
    • All IV antibiotics
    • Any antibiotic used to treat C.Diff
  – Laboratory
    • WBC
  – Microbiology
    • All dictated and transcribed culture results

• Data Fusion Engine
  – Mirth Data Bus
  – Receive and organize all data inputs
  – Output data to NRC and receive classified data
  – Privacy protection
    • Allow research while protecting against indirect identification
  – Data output in format suitable for analysis

• Analysis
  – STATA
  – DM Solutions (Mapping)
  – Data sets to DRDC, Carnegie Mellon (work in progress)
Results:

• Process Framework
  – Dealing explicitly with non technical factors is key to success.
  – Despite “friendly” environment, data access took much longer than expected.

• Technical Framework
  – Worked extremely well
  – Classification of unstructured data a key component
    • Equal collaboration between content experts and computer scientists.
CXR Infiltrate Syndrome

• Findings
  – Detected using NLP of Radiologist report
  – Include
    • Pneumonia, consolidation, cavitation – most specific
    • Infiltrate – somewhat specific
  – Exclude
    • Pleural effusion, increased markings

• Identify events likely to have occurred in hospital
  – Require previous negative CXR earlier in encounter
  – Eliminate events occurring first two days

• We did not differentiate between ventilated and not ventilated patients
Chest X Ray

Normal

Infiltrate
CXR Infiltrate as a Syndrome to Monitor Pneumonia

• Likely sensitive but not specific
• Potential Causes
  – Infection (Pneumonia)
  – Atelectasis (Collapse of part of lung)
  – Pulmonary embolism
• All medically important
• Many events preventable
Monitoring of Blood Infection

• Positive blood culture results identified
• Organisms Categorized by NLP
  – High Risk
  – Low Risk
• Syndromes
  – Any culture positive
  – High Risk
  – Low Risk
  – One high risk or two low risk
Results: UOHI Four Year Dataset

- University of Ottawa Heart Institute
  - April 1, 2008 – March 31, 2012

- Encounters (Admissions) – 21,452

- Unique Patients – 17,194

- Chest X Rays (CXRs) – 41,720
  - CXRs showing an infiltrate – 10,546
  - Encounters with any CXR infiltrate - 4,575
  - Encounters with new* CXR infiltrate - 2,266

- Positive Blood Cultures – 253
  - All on different encounters

*previous negative CXR that encounter and not on first 2 days of encounter.
CXR Infiltrate Syndromes vs. Hospital Mortality

All p-values < 0.000001

*2558 patients with pre-existing infiltrate excluded
New Infiltrate vs. Hospital Mortality
Sensitivity to LOS before Infiltrate Occurred

<table>
<thead>
<tr>
<th>Admitted at least ____ before infiltrate occurred</th>
<th>Mortality % (N)</th>
<th>Odds Ratio*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New Infiltrate</td>
<td>No Infiltrate**</td>
</tr>
<tr>
<td>2 days</td>
<td>6.17 (1831)</td>
<td>1.25 (9580)</td>
</tr>
<tr>
<td>3 days</td>
<td>6.86 (1472)</td>
<td>1.21 (8034)</td>
</tr>
<tr>
<td>4 days</td>
<td>6.81 (1218)</td>
<td>1.12 (6832)</td>
</tr>
<tr>
<td>5 days</td>
<td>6.55 (992)</td>
<td>1.19 (5648)</td>
</tr>
<tr>
<td>6 days</td>
<td>7.04 (810)</td>
<td>1.24 (4602)</td>
</tr>
<tr>
<td>7 days</td>
<td>7.34 (681)</td>
<td>1.44 (3664)</td>
</tr>
<tr>
<td>8 days</td>
<td>8.04 (560)</td>
<td>1.50 (2996)</td>
</tr>
<tr>
<td>9 days</td>
<td>8.54 (480)</td>
<td>1.41 (2484)</td>
</tr>
<tr>
<td>10 days</td>
<td>9.22 (412)</td>
<td>1.50 (2117)</td>
</tr>
</tbody>
</table>

**Patients with infiltrate on initial CXR excluded.

*All p values < 0.00001
New Infiltrate vs. Length of Stay (LOS)  
Sensitivity to LOS before Infiltrate Occurred

<table>
<thead>
<tr>
<th>Admitted at least ____ before infiltrate occurred</th>
<th>Mean Length of Stay Days (N)</th>
<th>Excess LOS* (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New Infiltrate</td>
<td>No Infiltrate**</td>
</tr>
<tr>
<td>2 days</td>
<td>21.4 (1831)</td>
<td>8.4 (9580)</td>
</tr>
<tr>
<td>3 days</td>
<td>23.4 (1472)</td>
<td>9.5 (8034)</td>
</tr>
<tr>
<td>4 days</td>
<td>25.1 (1218)</td>
<td>10.4 (6832)</td>
</tr>
<tr>
<td>5 days</td>
<td>26.9 (992)</td>
<td>11.6 (5648)</td>
</tr>
<tr>
<td>6 days</td>
<td>29.0 (810)</td>
<td>12.8 (4602)</td>
</tr>
<tr>
<td>7 days</td>
<td>30.6 (681)</td>
<td>14.3 (3664)</td>
</tr>
<tr>
<td>8 days</td>
<td>33.2 (560)</td>
<td>15.7 (2996)</td>
</tr>
<tr>
<td>9 days</td>
<td>40.0 (480)</td>
<td>17.1 (2484)</td>
</tr>
<tr>
<td>10 days</td>
<td>36.9 (412)</td>
<td>18.3 (2117)</td>
</tr>
</tbody>
</table>

**Patients with infiltrate on initial CXR excluded.  

All p values < 0.00001
EMR Based Surveillance of Blood Infections vs. Survival and Length of Stay (LOS)

### Hospital Mortality

<table>
<thead>
<tr>
<th>Mortality</th>
<th>+v Blood Culture</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>n</td>
<td>20,614</td>
<td>194</td>
</tr>
<tr>
<td>%</td>
<td>97.24</td>
<td>76.68</td>
</tr>
<tr>
<td>1</td>
<td>585</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>2.76</td>
<td>23.32</td>
</tr>
<tr>
<td>Total</td>
<td>21,199</td>
<td>253</td>
</tr>
</tbody>
</table>

### LOS prior to hospital death > day 2

<table>
<thead>
<tr>
<th>Group</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>Std. Dev.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>387</td>
<td>17.26804</td>
<td>1.038813</td>
<td>20.46224</td>
<td>15.22562 - 19.31046</td>
</tr>
<tr>
<td>1</td>
<td>55</td>
<td>45.81132</td>
<td>6.394281</td>
<td>45.55107</td>
<td>32.98026 - 58.64238</td>
</tr>
<tr>
<td>combined</td>
<td>441</td>
<td>20.69841</td>
<td>1.269484</td>
<td>26.65917</td>
<td>18.20341 - 23.19342</td>
</tr>
</tbody>
</table>

### LOS: All Patients surviving > day 2

Two-sample t test with equal variances

<table>
<thead>
<tr>
<th>Group</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>Std. Dev.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>14000</td>
<td>10.82879</td>
<td>0.969654</td>
<td>11.4731</td>
<td>10.63872 - 10.1088</td>
</tr>
<tr>
<td>1</td>
<td>245</td>
<td>40.92245</td>
<td>2.393672</td>
<td>37.4669</td>
<td>36.20755 - 45.6373</td>
</tr>
<tr>
<td>combined</td>
<td>14245</td>
<td>11.34637</td>
<td>1.00832</td>
<td>12.98937</td>
<td>11.13304 - 11.5596</td>
</tr>
</tbody>
</table>

### LOS in Survivors

<table>
<thead>
<tr>
<th>Group</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>Std. Dev.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2061</td>
<td>7.450907</td>
<td>0.699998</td>
<td>10.93735</td>
<td>7.313878 - 7.587936</td>
</tr>
<tr>
<td>1</td>
<td>194</td>
<td>39.05155</td>
<td>2.49094</td>
<td>34.69478</td>
<td>34.13859 - 43.96451</td>
</tr>
<tr>
<td>combined</td>
<td>2888</td>
<td>7.745531</td>
<td>0.766036</td>
<td>10.96355</td>
<td>7.596558 - 7.894504</td>
</tr>
</tbody>
</table>

\[ \text{diff} = \text{mean}(0) - \text{mean}(1) \]
\[ t = -37.7001 \]

All p-values < 0.0000001
Positive Blood Culture vs. Mortality

<table>
<thead>
<tr>
<th>In hospital at least</th>
<th>Hospital Mortality (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Risk</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>2 days</td>
<td>0.8</td>
</tr>
<tr>
<td>5 days</td>
<td>1.2</td>
</tr>
<tr>
<td>10 days</td>
<td>2.2</td>
</tr>
<tr>
<td>20 days</td>
<td>4.2</td>
</tr>
<tr>
<td>40 days</td>
<td>8.3</td>
</tr>
</tbody>
</table>

All p values < 0.00001

* ≥ one high risk or 2 low risk cultures
Positive Blood Culture vs. Total Length of Stay (LOS)

<table>
<thead>
<tr>
<th>In hospital at least</th>
<th>Length of Stay</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Any Positive</td>
<td>Low Risk</td>
<td>High Risk</td>
<td>Syndrome*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>2 days</td>
<td>10.8</td>
<td>40.9</td>
<td>11.0</td>
<td>43.3</td>
<td>11.1</td>
</tr>
<tr>
<td>5 days</td>
<td>14.1</td>
<td>42.5</td>
<td>14.3</td>
<td>45.0</td>
<td>14.4</td>
</tr>
<tr>
<td>10 days</td>
<td>21.4</td>
<td>46.6</td>
<td>21.8</td>
<td>49.1</td>
<td>22.0</td>
</tr>
<tr>
<td>20 days</td>
<td>35.4</td>
<td>59.3</td>
<td>36.1</td>
<td>61.4</td>
<td>36.6</td>
</tr>
<tr>
<td>40 days</td>
<td>63.5</td>
<td>78.0</td>
<td>64.4</td>
<td>80.4</td>
<td>65.1</td>
</tr>
</tbody>
</table>

All p values < 0.00001

* > one high risk or 2 low risk cultures
University of Ottawa Heart Institute
Rates of New Infiltrates April 2008 – March 2012 by Clinical Unit and Quarter (per 1000 bed-days)

1 = Q2-2008, 16 = Q1-2012
University of Ottawa Heart Institute Data Fusion

Q2-2008
April 8, 2008
Ontario Ministry of Health and Long Term Care:
Publicly reported infection rates by Hospital

Ventilator Acquired Pneumonia

Central Line Infection
MOHLTC Reported Rates

Ventilator Acquired Pneumonia

Central Line Infection

EMR Information Fusion

New CXR Infiltrates

Positive Blood Cultures
Next Steps

• Methods Development
  – Surveillance of unexpected events
  – Pre-event identification of high risk patients
  – Covariates: syndrome definition and development

• Potential Projects
  – On line EMR based surveillance of hospital acquired infections
  – Surveillance of reportable infectious diseases via OLIS
  – Surveillance of high risk respiratory syndromes using CXR results
  – Populate diabetes registry using unstructured EMR data
EMR-Based Diabetes Surveillance

• Problem
  – Clinically relevant data in EMR is unstructured.
  – Much exists as free text
  – Inaccessible for population-based analytics

• Approach
  – Structure unstructured data, targeting diabetes and specific parameters relevant to treatment guidelines
    • ASA
    • ACE
    • Blood Pressure
    • Blood Sugar
    • Cholesterol
  – NLP algorithms to extract relevant information directly from free-text fields in EMR

• Development test bed
  – 80,000 dictated and transcribed clinical consult and follow up letters on 30,049 cardiac patients seen at UOHI
Results: Diabetes Report Card

Number of Patients with Diabetes

- Diabetic (23.04%)
- Non-Diabetic (76.96%)

Patients on Ramipril

<table>
<thead>
<tr>
<th>Number of Patients with Diabetes</th>
<th>Diabetic</th>
<th>Non-Diabetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5000</td>
<td>15000</td>
</tr>
<tr>
<td>5</td>
<td>2000</td>
<td>12000</td>
</tr>
<tr>
<td>10</td>
<td>1500</td>
<td>8000</td>
</tr>
<tr>
<td>15</td>
<td>1000</td>
<td>4000</td>
</tr>
<tr>
<td>20</td>
<td>500</td>
<td>1000</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>500</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>20</td>
</tr>
</tbody>
</table>

Patient Info

- Patient ID: 10000001
- Diabetic: Yes
- YOB: 1963
- Gender: Female
- Age (last visit): 46

LAB Data

- Blood Pressure: May 13, 2009 140/80
- Creatinine: May 13, 2009 92
- Fasting Glucose: May 13, 2009 6.4
- HDL: Oct 08, 2008 1.06
- LDL: Oct 08, 2008 2.5
- Weight: May 13, 2009 155

Medication

- Entrophen: May 13, 2009 81 mg
- Metformin: May 13, 2009 500 mg
- Pravachol: May 13, 2009 40 mg
- Vasotec: May 13, 2009 10 mg

Measure | Grade
---------|--------
ACE      | 1      
ARB      | 0      
Blood Pressure | 0.5 
LDL      | 0.5    
HbA1C    | 0      
Fasting Glucose | 1    
BMI      | 0      
Total    | 3/7    

Diabetes Report Card

University of Ottawa Heart Institute/The Ottawa Hospital
Conclusion

- Disease Surveillance using EMR data is possible.
- Advantages
  - Potential access to very large amounts of data
  - Methods, once developed, are scalable
  - Objective data directly from providers
- Key components
  - Process Framework
  - Engagement of Subject Matter Experts
  - Database management
  - Data Science
  - Statistics and visualization
- Success factors
  - SME involvement
  - Development of Data Science
  - Attention to process and non-technical factors