Patient Identification and Matching – Fundamental to a National Health Information Network

Testimony to NCVHS, Standards and Security Subcommittee

Scott Schumacher, Ph. D
Lorraine Fernandes, RHIA

September 21, 2005
Patient identification and patient matching

Overview:

<table>
<thead>
<tr>
<th>Icon</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Computer" /> <img src="image" alt="Patient" /></td>
<td>Patient identification technology is proven and widely used today to create an EHR</td>
</tr>
<tr>
<td><img src="image" alt="ROI" /></td>
<td>Managing patient identities across the healthcare ecosystems requires a flexible, interoperable architecture that adapts to varying standards</td>
</tr>
<tr>
<td><img src="image" alt="Currency" /> <img src="image" alt="ROI" /></td>
<td>A new, expensive national healthcare identifier is not needed</td>
</tr>
<tr>
<td><img src="image" alt="Eye" /> <img src="image" alt="Lock" /></td>
<td>The federated approach to patient identification and building an EHR is the best way to manage accuracy, privacy and security</td>
</tr>
<tr>
<td><img src="image" alt="Plus" /> <img src="image" alt="Minus" /></td>
<td>Algorithms for patient identification must balance false positives and false negatives by using ratios of likelihood and probabilities of weight/scoring</td>
</tr>
<tr>
<td><img src="image" alt="Map" /></td>
<td>Canada uses patient identification technology to facilitate provincial and national views</td>
</tr>
<tr>
<td><img src="image" alt="Map" /></td>
<td>The proven, federated architecture to manage regional, state, and national patient matching is available today and can be deployed effectively in the U.S.A.</td>
</tr>
</tbody>
</table>
Person identification technology is widely used today to create EHR, RHIO, and NHIN

- Offices Chicago, CA, Austin, Phoenix
- Over 2 billion records analyzed
- 1400 healthcare facilities use Initiate technology
- Typically discover duplication rates of 15-30% in “clean” files
- Installations from 500K over 500 million records
- Search and link across 150 million records in under ¼ of a second

... Establish patient identity

... Prescribe a drug

... Create national prototype

... Process a claim

... Access clinical info on demand

... 360° view of pharmacy customers

... Share data securely

... Create a national EHR

... Nationwide physician search
Patient identification enables tomorrow’s virtual health record

EHRs, RHIOs and NHIN

- Improve patient care and reduce medical risks
- Improve efficiency by reducing redundant care activities
- Support consumer directed health information management
- Comply with regulations
- Enhance operational productivity and efficiency

ROI
National health ID: No “magic bullet”

- Just another piece of data
  - As likely to have errors as existing methods
- Long and expensive process
  - Hard to implement locally, almost impossible nationally
  - Hard to drive adoption in existing IT systems
  - Few benefits from partial implementation
- Political culture of the US not amenable to national identifiers
- Need to link this ID to several billion existing medical records
- Risk of privacy spills significantly worsened with universal identifier
- Discussed by Connecting for Health but recommended federated, probabilistic approach
Federated patient ID manages privacy & security

EHR components
- Public Health
- Physician
- Mental Health
- IDN
- Reference & Referral Lab Centers
- Pharmacy
- Hospital

Search criteria -
<table>
<thead>
<tr>
<th>Last</th>
<th>First</th>
<th>Middle</th>
<th>DOB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson</td>
<td>Robert</td>
<td></td>
<td>5/21/1960</td>
</tr>
</tbody>
</table>

Virtual EHR - Robert E. Johnson

<table>
<thead>
<tr>
<th>Facility</th>
<th>Local MRN</th>
<th>Date of Last Service</th>
<th>Last</th>
<th>First</th>
<th>Middle</th>
<th>DOB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital</td>
<td>A93452</td>
<td>12/24/2004</td>
<td>Johnson</td>
<td>Bob</td>
<td>E</td>
<td>5/21/1960</td>
</tr>
<tr>
<td>IDN</td>
<td>583959</td>
<td>7/15/2001</td>
<td>Johnson</td>
<td>Robert</td>
<td>E</td>
<td>5/12/1960</td>
</tr>
<tr>
<td>Reference Lab</td>
<td>128455</td>
<td>8/30/1998</td>
<td>Johnson</td>
<td>Robert</td>
<td></td>
<td>5/21/1960</td>
</tr>
</tbody>
</table>

Initiate Identity Hub™

ER Visit Allergic reaction
Federated approach provides common ground for the privacy concerns

- Inaccurate data leads to false positives, inflexible model makes correcting mistakes difficult
- Too much data being shared
- Need to avoid unique identifier
- Need ability to audit - who is accessing data and when

- Need for access to large amounts of data in real-time, stored in heterogeneous environments
- Need quickly deployed, non-intrusive solution
Patient matching in a federated model

- Initiate has created an identity engine which facilitates fast, efficient record matching based on demographic information

- The Initiate Identity Hub™ technology is populated with the optimal algorithms for matching on this information

- The engine and algorithms scale to any problem

<table>
<thead>
<tr>
<th>Name</th>
<th>CC#</th>
<th>Address</th>
<th>City</th>
<th>State</th>
<th>Zip</th>
<th>Phone</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kate Lamb</td>
<td>cc#5555-55-1234</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="mailto:klamb@aol.com">klamb@aol.com</a></td>
</tr>
<tr>
<td>Mrs. K. Jones</td>
<td></td>
<td>1000 Main St.</td>
<td>Hertowne</td>
<td>CA</td>
<td>45883-2539</td>
<td>(555) 132-4567</td>
<td></td>
</tr>
<tr>
<td>Katherine J. Jones</td>
<td>cc#5555-66-1234</td>
<td>Hertown</td>
<td>CO</td>
<td>55883</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catherine Lamb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(555) 123-4567</td>
<td><a href="mailto:klamb@aol.com">klamb@aol.com</a></td>
</tr>
<tr>
<td>Katherine J. Jones</td>
<td>cc#5555-55-1234</td>
<td>1000 Main St.</td>
<td>Hertown</td>
<td>CA</td>
<td>45883-2539</td>
<td>(555) 123-4567</td>
<td><a href="mailto:klamb@aol.com">klamb@aol.com</a></td>
</tr>
</tbody>
</table>
Initiate Identity Hub™ matching algorithm

- Configurable
- Decision theoretic basis – uses likelihood ratio test to determine if two records refer to the same object or not
- Test statistic comprises contribution from individual attributes
  - Comparison techniques specific to attribute types which are robust to typical errors based upon data experience. (e.g. Name comparison considers phonetic spelling, position misalignment, initials vs. name, etc.)
  - Comparison techniques which are general and can be applied to arbitrary attributes
  - These techniques are applied to available attributes to create final test statistic
- Underlying probability densities for the test are estimated from the data
Addressing false negatives (missed searches)

- False negatives typically caused by
  - Variation in recording demographic information – Use of nicknames, misspellings, name reversals, etc.
  - Missing or invalid attributes (e.g. No DOB)

- To combat variation, the algorithm requires a robust set of comparison routines
  - e.g. for names, Initiate considers 1) exact match, 2) nickname match, 3) phonetic match, and 4) name to initial matches. We also test all possible name token alignments
  - For SSN we look for common typographical errors
  - Important to address these in candidate selection as well

- When addressing “thin” data, making the best use of the data you do have becomes critical
  - Probabilistic scoring based upon observed data is key
Addressing false positives (false returns)

- False positives typically caused by
  - Matches on commonly occurring attribute values
  - Ad hoc combination of attribute scores
  - Multiple members from the same family

- Weighting matches based upon observed frequencies address commonly occurring attributes
  - We use a probabilistic scoring based upon analysis of client data
  - Employ a likelihood ratio test which weights the match contribution of individual attributes naturally
  - Family members are treated via a post-detection algorithm

- Scalability is a key issue – ad hoc weighting schemes typically don’t scale to large files sizes
Weights/scoring

- Given a set of attribute matching outcomes how do you decide if the records refer to the same entity or different entities?
- Need to look at ratios of probabilities
  - What are the probabilities of these outcomes if you know that the records referred to the same entity?
  - What are the probabilities of these outcomes if you know that the records refer to different entities?
- Weights
  - Match weights are essentially determined by knowing the uniqueness of the attribute value
  - Mismatch weights are determined by knowing how often an attribute is entered correctly
  - These probabilities are determined from analyzing the data file
Impact of data quality

- **Analytical simulation of matching performance**
  - Single threshold – low false-positive rate
  - Search against 10 million member database
- **Four attributes - name, DOB, Zip, SSN**
- **Vary data quality**
  - Fraction of the time an attribute is available
  - Full SSN or only the last 4-digits
- **Simulate false-negative rate**

<table>
<thead>
<tr>
<th>Name</th>
<th>DOB</th>
<th>Zip</th>
<th>SSN</th>
<th>False-negative rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>0%</td>
<td>6%</td>
</tr>
<tr>
<td>100%</td>
<td>90%</td>
<td>90%</td>
<td>0%</td>
<td>22%</td>
</tr>
<tr>
<td>100%</td>
<td>90%</td>
<td>90%</td>
<td>70%</td>
<td>7%</td>
</tr>
<tr>
<td>100%</td>
<td>90%</td>
<td>90%</td>
<td>70% (4-digits)</td>
<td>8%</td>
</tr>
</tbody>
</table>
Canada’s proven, federated model of patient identification

Six Provinces use Initiate Identity Hub™ software
National view architecture model

- Regional or state hubs with peer-to-peer communications for sharing and retrieving patient information
National Health Information Network

Mission Statement: To create an interconnected, electronic health information infrastructure to support patient safety and better healthcare
### Patient identification and patient matching

<table>
<thead>
<tr>
<th>Icon</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>🧁</td>
<td>Patient identification technology is proven and widely used today to create an EHR</td>
</tr>
<tr>
<td>📢</td>
<td>Managing patient identities across the healthcare ecosystems requires a flexible, interoperable architecture that adapts to varying standards</td>
</tr>
<tr>
<td>💰</td>
<td>A new, expensive national healthcare identifier is not needed</td>
</tr>
<tr>
<td>👀🔒</td>
<td>The federated approach to patient identification and building an EHR is the best way to manage accuracy, privacy and security</td>
</tr>
<tr>
<td>🔍❌</td>
<td>Algorithms for patient identification must balance false positives and false negatives by using ratios of likelihood and probabilities of weight/scoring</td>
</tr>
<tr>
<td>🇨🇦</td>
<td>Canada uses patient identification technology to facilitate provincial and national views</td>
</tr>
<tr>
<td>🇺🇸</td>
<td>The proven, federated architecture to manage regional, state, and national patient matching is available today and can be deployed effectively in the U.S.A.</td>
</tr>
</tbody>
</table>