Identity Is Key: How to Unlock Big Data and Analyze Populations

Brent Williams
Session Objectives

• Provide the audience with an understanding of how aggregating information from multiple sources (claims, clinical, biographical, etc.) is the key to analyzing populations

• Discuss how the challenge of aggregating information boils down to the problem of correctly attributing this information to the correct patient

• Inform the audience that identity information is the key to correctly attributing information to the correct patients, because current “matching” technologies rely on identity attributes like name, address, and birthdate to match information to patients

• Go over how a big data approach to identity information enables more successful and accurate information attribution
Why is population health important?

- Raise Incentives
- Lower Costs
- Patient Health and Welfare
- More Timely
Population health: the “Old World”

Minimal analysis can be performed
Population health: the “Current World”

Course of care

More detailed analysis to provide better outcomes in the future
Population health: the “Emerging World”

More detailed analysis to provide better outcomes in the future
Population health: the “Target World”

More detailed analysis of past and present to provide better outcomes in the future
Imagine if my providers had a 360-degree view of me
How do you aggregate and access this data?

- **CLAIMS**
  - Diagnoses
  - Prescriptions
  - Care events

- **CLINICAL**
  - Lab results
  - Radiology

- **BIOGRAPHIC**
  - Family
  - Location
  - Employment

- **OTHER**
  - Non-medical events
  - Connected medical devices

Link and exchange patient information across systems and organizations.

Enrich patient information with third-party data.
Gain longitudinal views of patients to better analyze and understand populations.

Analytics that are:
- Deeper
- More complete
- More accurate
Gain more relevant courses of care and more granular outcome measurements

Analytics that are:
• Deeper
• More complete
• More accurate
Gain more relevant courses of care and more granular outcome measurements

Analytics that are:
- Deeper
- More complete
- More accurate
Apply the analytics in the clinical or aftercare environment

Analytics that are:
- Deeper
- More complete
- More accurate

History of poor diet and exercise

30% higher risk for heart disease

Solicit support of family to develop new regimen

New prescription

Dust exposure

Lung function test
Identity data is the key to linking patient records and appending big data

- CLAIMS
  - Diagnoses
  - Prescriptions
  - Care events

- CLINICAL
  - Lab results
  - Radiology

- BIOGRAPHIC
  - Family
  - Location
  - Employment

- OTHER
  - Non-medical events
  - Connected medical devices

Name
Address
Date of Birth
SSN
Gender
Identity data is a collection of attributes that often change over time

**Bio graph**
- Name(s), Address, Gender, Birthdate

**Family graph**
- Parents, Siblings, Spouses, Children

**Cyber graph**
- Phone, Email, Social IDs, Device IDs, IP

**Government graph**
- SSN, National Benefits ID, Drivers’ Permits, Legal Records

**Social graph**
- Friends, Roommates, Associations

**Professional graph**
- University, Degrees, Employers, Coworkers, Prof credentials, Associations

**Commercial graph**
- Loyalty cards, Purchases

**Financial graph**
- Fin. Accounts, Homes, Cars, Boats, Mortgages, Loans, Credit
30-40% of all medical records have errors in their identity data

<table>
<thead>
<tr>
<th>One person’s data</th>
<th>... experiences a wide array of errors ...</th>
<th>... resulting in disparate identity representations across systems</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Name change</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Address change</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phone/email change</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Married</td>
<td>Divorced</td>
</tr>
<tr>
<td></td>
<td>Hyphenated name</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nicknames</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jr/Sr, Twins</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Default entries</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Missing data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Errors</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Spelling error</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transcription error</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Homonym error</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data Gov.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Privacy controls</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Formatting</td>
<td></td>
</tr>
</tbody>
</table>
There is a new way to match patient identities together that goes beyond deterministic or probabilistic matching algorithms to make a match despite bad identity data.

It’s called referential matching and it utilizes big data.
Referential matching utilizes big data databases of reference identity data.

These big data databases aggregate identity data from various sources:

- Credit header data
- Telco record identity data
- Gov’t & legal record identity data
Two patient records with very different data can still match to each other.
Two patient records with very different data can still match to each other.
Better matching using a big data approach enables the future of population health.
Brent Williams
Brent.Williams@verato.com
Twitter: @jbrentwilliams
linkedin.com/in/jbrentwilliams