Considerations for Development and Use of a Master Person Index (MPI)

July 26, 2016
3 - 4 pm EST
Presenters

Clare Tanner, PhD  
Co-Director of Data Across Sectors for Health (DASH),

Melissa Moorehead  
Policy Analyst and Project Manager, Michigan Public Health Institute

Stephen Singer, MCP  
Senior Manager of Data Analytics, Camden Coalition of Healthcare Providers

Dan Chavez, MBA  
Executive Director, San Diego Health Connect
Meeting Information

▪ Meeting Link:
  http://academyhealth.adobeconnect.com/mpi/

▪ Registered: Select “Enter with your login and password” and enter the following:
  ▪ Username: [enter email address used to register for the webinar]
  ▪ Password: index
  ▪ Click “Enter Room”

▪ Unregistered Guest: Select “Enter as a guest” and enter your name, e.g., Kelsi Feltz, CHP.
Meeting Information, continued

▪ Conference Line: 1-866-546-3377

▪ Access Code: 6478553818

▪ Reminders:
  ▪ Please hard-mute your computer speakers and the speakers in the web conference
  ▪ Please mute your phone line when you are not speaking to minimize background noise

▪ Technical difficulties? Email us at chpinfo@academyhealth.org
Chat Feature

- To share your comments using the chat feature:
  - Click in the chat box on the left side of your screen
  - Type into the dialog box and click the send button

- To signal to presenters you have a question / comment:
  - Click on the drop down menu near the person icon and choose *raise your hand*
Agenda

- **Introduction (3 minutes)**
  - Clare Tanner, DASH NPO, will provide a brief introduction to All In

- **MPI Case Study #1 (12 minutes)**
  - Stephen Singer, MCP, Senior Manager of Data Analytics at Camden Coalition of Healthcare Providers, will discuss how Camden Coalition uses and continues to evolve their person-level matching using various methodologies in the research settings.

- **MPI Case Study #2 (12 minutes)**
  - Daniel Chavez, MBA, Executive Director at San Diego Health Connect and a CHP Subject Matter Expert community, will discuss how San Diego Health Connect is using an HIE and addressing standards to improve automated patient matching capability.

- **Discussion (30 minutes)**

- **Wrap-Up (3 minutes)**
DASH and CHP are All In!

Community Health Peer Learning Program

- NPO: AcademyHealth, Washington D.C.
- Funded by the federal ONC
- 15 participant and subject matter expertise communities

Data Across Sectors for Health (DASH)

- NPO: Illinois Public Health Institute in partnership with the Michigan Public Health Institute
- Funded by the RWJF
- 10 grantee communities
All In: Data for Community Health

1. Support a movement acknowledging the social determinants of health

2. Build an evidence base for the field of multi-sector data integration to improve health

3. Utilize the power of peer learning and collaboration
Considerations & Questions about Record Linkage & MPI’s

Stephen Singer, Senior Program Manager, Data Analytics & Quality Improvement

Camden Coalition of Healthcare Providers
The Camden Coalition Data Environment

**HIE**
- vendor-managed.
- **MPI via** ...
  - a black box
- IDs & events
- corrections

**user-customizable,**
- vendor-hosted.
- **MPI via HIE linkage**
- + deterministic linkage
- + extensive manual review

Internal performance & care tracking

**home-grown PostgreSQL database.**

**No MPI.** previously linked via commercial probabilistic linkage software, temporarily via hierarchical, fuzzy, deterministic match

**me**
- retrospective hospital claims
- cross-sector integrated data
Our HIE

Hospitals
- Virtua
- Lourdes Health System
- Cooper University Health Care
- Kennedy Health

Labs
- Quest Diagnostics
- LabCorp
- BioReference Laboratories

Perinatal Risk Assessments
- Southern New Jersey Perinatal Cooperative

MCOs
- UnitedHealthcare Community Plan
- Horizon

Primary Care Practices
- Cooper Family Medicine
- Cooper Urban Health Institute
- Fairview Village Family Practice
- Lourdes Medical Associates
- Ramon Acosta, PC
- Reliance Medical Group
- Virtua Kyle Will Family Health Center

FQHCs
- CAMcare
- Project HOPE

Other Community Partners
- CFG Health Systems
- Holy Redeemer Home and Hospice
- LIFE at Lourdes
- Northgate II
- Oaks Integrated Care
- Planned Parenthood
- South Jersey Behavioral Health Resources

State of New Jersey Department of Human Services Division of Medical Assistance & Health Services

Our HIE

Camden Coalition

HEALTH INFORMATION EXCHANGE

- Master Patient Index
- Admission / Discharge / Transfer (ADT)
- Lab/Radiology Results
- Discharge Summaries
- Medications
- Claims Data
- Perinatal Risk Assessments

Contribute

Read/Write

Claims Data
Cross-sector Integrated Data “System”

Existing Data Sharing:
1) All-payer hospital claims from 4 regional health systems biannual (plus a 1 time extract from a 5th)
2) State Medicaid Claims monthly
3) Camden Police Department no fixed schedule (arrest, call-for-service, & overdose)
4) Camden City School District no fixed schedule (enrollment, truancy, absenteeism, & suspension data)
5) Camden County Jail (booking & release) monthly
6) NJ State Prison (booking & release) bi-monthly
7) property data (citywide vacancy survey) one time

In Discussion:
1) Homelessness Management Information System
2) State Mortality Records
To resolve existing data dis-integration \((\text{linkage})\) & prevent future data dis-integration \((\text{data management})\)

So that we can correctly identify & characterize patients for appropriate & coordinated care, accurate quality metrics, and research?
Some data are undecidably ambiguous. (What about twins?)
New data require unstable IDs.
Data entry is only partially controllable.
Data entry isn’t the only source of error.
1. How soon? How fast?
2. How expensive?
   ($ + \textit{training} + \textit{staff-hours})
3. How flexible & stable?
4. How interoperable?
5. How accountable?
5 questions for any vendor:

1. Can I get ALL of my data back?
2. How do you do it?
3. Who can I talk to... outside of sales and marketing?
4. How responsive is tech support?
5. Can you flag records by linkage quality?
<table>
<thead>
<tr>
<th>hospital</th>
<th>mrn</th>
<th>dob</th>
<th>last</th>
<th>first</th>
<th>mid</th>
<th>ssn</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>1</td>
<td>06/03/1965</td>
<td>SMITH</td>
<td>SIMON</td>
<td></td>
<td>296 1 4 6 4 1 1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>HIGHSMITH</td>
<td>BEN</td>
<td></td>
<td>296 1 4 6 5 1 1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>RUIZ</td>
<td>SIMON</td>
<td></td>
<td>296 1 4 6 4 1 1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>RUIZ</td>
<td>BENN</td>
<td></td>
<td>296 1 4 6 5 1 1</td>
</tr>
<tr>
<td>X</td>
<td>5</td>
<td>06/20/1965</td>
<td>SMITH</td>
<td>LARRY</td>
<td>N</td>
<td>296 1 4 6 5 1 0</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>296 1 3 6 5 5 5</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
<td>RUIZ</td>
<td>BEN</td>
<td>L</td>
<td>296 1 4 6 5 1 0</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>296 1 4 6 5 1 0</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td></td>
<td>RUIZ</td>
<td>SYMON</td>
<td></td>
<td>296 1 4 5 6 1 0</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>296 1 4 6 5 1 0</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>06/20/1966</td>
<td>SMITH</td>
<td>LARRY</td>
<td>N</td>
<td>296 1 3 6 5 5 5</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>06/30/1966</td>
<td></td>
<td>JAMES</td>
<td></td>
<td>296 1 4 5 8 8 8</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>296 1 4 6 4 1 1</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td></td>
<td>RUIZ</td>
<td>BEN</td>
<td></td>
<td>296 1 4 5 5 6 0</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>296 1 7 6 4 1 1</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td></td>
<td>RUIZ-SMITH</td>
<td>LARRY</td>
<td>N</td>
<td>296 1 4 6 4 1 1</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td></td>
<td>SMITH</td>
<td></td>
<td></td>
<td>296 1 4 6 4 1 1</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>296 1 4 6 4 1 5</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>296 1 4 6 4 1 1</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>296 9 8 6 8 4 3</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>296 1 4 6 4 1 1</td>
</tr>
</tbody>
</table>
Data Production (out of your hands)

- Design constraints
- Documentation error
- Noise & obfuscation
- Clerical error
- Extract error
- Data structure mismatch
- Linkage error
- Contamination
- Technical glitch
- Omission
- Unimplemented feature

Errors, errors, everywhere!

Data Manipulation

- Manual Entry
- Hospital 1 Database
- Business Rules
- Record Linkage
- Merged Data
- Processing
- Analysis files
### Deterministic linkage

Deterministic linkage groups together records that are equal on subsets of identifier fields.
Probabilistic linkage calculates a total score for two records to determine how likely it is that both refer to the same individual. The total score is the sum of scores generated by the comparison of individually weighted fields.
Bursting the Linkage Bubble

1. Probabilistic is better *when assumptions hold*
2. Linkage success depends on geography, ethnicity, poverty, and other health-correlated variables.
3. String comparators make a bigger difference than other tweaks to linkage methods.
4. ~80% of the effort and improvement is not even in the linkage method, it’s in data cleaning and preparation, but you can over-clean *and* under-clean!
What else would you like to discuss?

Name parsing

Twins

String comparators

Phonetic algorithms

SSN’s

Other data cleaning processes, terms & issues

probabilistic linkage software

using graph databases to manage linking data

request process for external data

Etc.!
Patient Records Matching
Overcoming the largest obstacle to health information exchange: One HIE’s story

Daniel Chavez, Executive Director
San Diego Health Connect
The SDHC mission

Our Mission
To connect healthcare stakeholders to deliver quality, comprehensive information for better care.

When every individual’s health information is securely available to their doctors when and where they need it:

- Doctors can provide better, more informed care.
- Duplication of tests and procedure decreases.
- Costs go down.
Participating organizations

UC San Diego Health System

Kaiser Permanente

SHARP

Rady Children's Hospital San Diego

Health and Human Services

Scripps

Health Quality Partners of Southern California
Trusted health information exchange...

- Is built on technical interoperability
  - HL7
  - FHIR
  - ISO
- Uses document standards to achieve functional interoperability
  - CCR
  - SNOMED
  - DICOM
  - LOINC
  - NCPDP
  - RxNorm
  - CPT
  - ICD-9/10
- Is enabled by semantic interoperability
  - Patient Matching:
    - No false positives
    - Minimal false negatives
SDHC uses an MPI as a record locator service.
When records do not match, records ended up in an “Exception queue”
Our working group decided we needed a better way to match records.

- Total members: 41
- Different organizations: 13
- Meetings per month: 2
Referential matching is a revolutionary new way to match patient records.

MPI matching (deterministic or probabilistic) can’t see through different or bad identity data.

Referential matching works despite different or bad identity data.

X No Match
How Verato is different

**Verato**

- Algorithm RICH
- Data RICH

**MPI Technologies**
- Algorithm RICH
- Data poor
- Probabilistic Matching
- IBM | Informatica | Oracle | SAP

**Information Providers**
- Data RICH
- Algorithm poor
- Fragmented Data
- Acxiom
- TransUnion
- Experian | Equifax

**Cloud-based**

Referential Matching + Complete, Unified Data
CARBON™ – the most comprehensive reference database of identities in the US
In total, SDHC increased the number of matches in its MPI by 110%
Futures – Improve edge case matching

1. Incorporate relationship data in ADTs

2. Pediatrics: add twins identifiers to patient data model at institutions

3. Develop twin inference algorithm for newborns to support twin analysis for adults
Futures – Accommodating varying data governance models

1. Understanding an organization’s identity data governance model

2. Demonstrating proof for non-obvious matches while maximizing privacy

3. Accommodating variations in transport protocols
Futures – Connect the Community

ConnectWellSD
Connect · Collaborate · Empower

Connecting All for Better Health & Wellness

COMMUNITY INFORMATION EXCHANGE

SAN DIEGO HEALTH CONNECT
Better Information • Better Care

2-1-1 SAN DIEGO
Futures – Connect the eHealth Exchange

HIOs and other Communities
Kaiser
SDHC
State / Local Governments
Academic Medical Centers
SSA VA / DOD CDC
“Better is possible. It does not take genius. It takes diligence. It takes moral clarity. It takes ingenuity. And above all, it takes a willingness to try.”

Atul Gawande
Questions?
Presenters

Clare Tanner, PhD
Co-Director of Data Across Sectors for Health (DASH),

Melissa Moorehead
Policy Analyst and Project Manager, Michigan Public Health Institute

Stephen Singer, MCP
Senior Manager of Data Analytics, Camden Coalition of Healthcare Providers

Dan Chavez, MBA
Executive Director, San Diego Health Connect
Connect with Us!

- Sign up for news from All In at dashconnect.org
- Follow us at @DASH_connect and @AcademyHealth at #CHPHHealthIT
- Contact information for speakers
  - Stephen Singer, stephen@camdenhealth.org
  - Dan Chavez, dchavez@sdhealthconnect.org
- Evaluation
- A resource list, slides, and recording will be available