Enhancing State-Based Data with Federally-Available Data: Linking Massachusetts Hospital Discharge Data to ResDAC

Zi Zhang, MD, MPH
Catherine Nwachukwu, MPH
Huong Trieu, PhD
Mark Paskowsky, MPP

NAHDO Annual Conference 2020
Agenda

- Background
- Methods
- Next Steps
- Questions & Answers
Background

- State and local health data organizations have widely used state-specific hospital discharge records and available claims data to analyze and report on health system performance.

- In Massachusetts, discharge summary records from inpatient, observation, and emergency department visits are used to measure key performance measures, such as hospital readmissions and revisits.

- Limitations of discharge summary records:
  - Expected payer source(s) and potential miscoding
  - Determination of primary vs. secondary payer source(s)
  - Unreliable or incomplete charge information
  - Limited or no patient enrollment information
Social Security Number is increasingly missing from acute care hospital discharge data.
Issues with linking patients by Social Security Number between Case Mix (acute care hospital data) and ResDAC:

- 82% of patients with Medicare indicated as their payer (based on MA acute care hospital case mix data) matched a ResDAC Medicare beneficiary

  **Note:** expected about 95+% 

- 59% of patients, aged 65+, who did not indicate Medicare as their payer (based on MA acute care hospital data) were actually found in the ResDAC Medicare data

  **Note:** expected low %
Background – CHIA’s Linkage Project

To address these challenges, the Massachusetts Center for Health Information and Analysis (CHIA) has undertaken a three-phase linkage process:

**Phase 1: Complete**
- Link patients within the Massachusetts Acute Hospital Case Mix Databases (Case Mix), including inpatient, observation, and emergency department data

**Phase 2: Near Completion**
- Link Case Mix to the Massachusetts All Payer Claims Database (APCD)

**Phase 3: Beginning Soon**
- Link the federally-available ResDAC data to the linked Case Mix-APCD data from Phase 2

*Delayed due to COVID-19*
Background – CHIA’s Linkage Project Phase 1

Trend in Discharges and Readmissions by Payer Type using EPI vs. SSN
SFY 2011-2018

Legend:
- Eligible Discharges (EPI)
- Eligible Discharges (SSN)
- Eligible Readmissions (EPI)
- Eligible Readmissions (SSN)
- Missing or Invalid SSN
Methods

- **Goal:** To create a Master Patient Index (MPI) by linking individuals across data sources (Case Mix, APCD, and ResDAC) and across time.

- **Linking Fields:** Social Security Number, date of birth, gender, zip code, name and address (if available), and other internal patient IDs such as medical record number, health plan subscriber ID, beneficiary ID.

- **Steps:**
  1. Begin with the APCD-Case Mix MPI data.
  2. Adapt the APCD-Case Mix probabilistic matching algorithm from Phase 2 by developing a new scoring matrix that satisfies all three data sources’ requirements for linking individuals.
  3. Add ResDAC Medicare Beneficiary Summary Data and run the algorithm to match patients across all three databases.
Methods – MPI Data Elements

The combined MDM algorithm will use these elements from the Case Mix, APCD eligibility, and ResDAC data:

<table>
<thead>
<tr>
<th>Case Mix Elements</th>
<th>APCD Elements</th>
<th>ResDAC Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient SSN</td>
<td>Member SSN</td>
<td>Beneficiary SSN</td>
</tr>
<tr>
<td>Patient Date of Birth</td>
<td>Member Date of Birth</td>
<td>Beneficiary Date of Birth</td>
</tr>
<tr>
<td>Patient Gender</td>
<td>Member Gender</td>
<td>Beneficiary Gender</td>
</tr>
<tr>
<td>Patient Zip Code</td>
<td>Member Zip Code</td>
<td>Beneficiary Zip Code</td>
</tr>
<tr>
<td>Patient Healthplan ID (from membership card)</td>
<td>Member Healthplan ID (Carrier-specific member ID)</td>
<td>Beneficiary Healthplan ID (Beneficiary ID)</td>
</tr>
<tr>
<td>Patient First Name</td>
<td>Member First Name</td>
<td></td>
</tr>
<tr>
<td>Patient Last Name</td>
<td>Member Last Name</td>
<td></td>
</tr>
<tr>
<td>Patient Address</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Enhancing State-Based Data with Federally Available Data | NAHDO 2020
First, patients are matched deterministically within data sources:

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Match Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case Mix</td>
<td>Records with the same OrgID and Medical Record Number</td>
</tr>
<tr>
<td>APCD</td>
<td>Records with the same OrgID and Carrier-Specific Member ID (CSUMID)</td>
</tr>
<tr>
<td>ResDAC</td>
<td>Records with the same Beneficiary ID</td>
</tr>
</tbody>
</table>
Methods – Match Rules (2)

Then, all other data elements are used to probabilistically match records:

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All elements agree.</td>
</tr>
<tr>
<td>2</td>
<td><strong>Any single</strong> element disagrees/is missing, all other elements agree.</td>
</tr>
<tr>
<td>3</td>
<td>SSN and DOB agree, all other elements agree but any two elements (not NAME) disagree.</td>
</tr>
<tr>
<td>4</td>
<td>DOB missing, all others agree but any one element (not SSN) disagrees.</td>
</tr>
<tr>
<td>5</td>
<td>SSN and DOB are missing, all other elements agree.</td>
</tr>
<tr>
<td>6</td>
<td>SSN, DOB, and HealthPlanID agree, all other elements are missing.</td>
</tr>
</tbody>
</table>
Methods – Special Considerations

Some insights we learned from Phase 2 Linkage:

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Different First Names, DOB missing, all others match</td>
<td>Could be same gender twins were born, given mom's SSN, DOB missing</td>
</tr>
<tr>
<td>2</td>
<td>Different Last Names, SSN missing, all others match</td>
<td>Without SSN present can't tell if LN change is the same person</td>
</tr>
<tr>
<td>3</td>
<td>Different Last Names, SSN disagrees, all others match</td>
<td>If SSN disagrees, can't tell if LN change is the same person</td>
</tr>
<tr>
<td>4</td>
<td>Different First and Last Names, all others match</td>
<td>Two different names, same town, same gender, same DOB, SSN transposed</td>
</tr>
<tr>
<td>5</td>
<td>Different First and Last Names and addresses, all others match</td>
<td>Not enough to tell if same person, transposed SSN in a diff. location</td>
</tr>
</tbody>
</table>
Methods – Linkage Quality Checks

- Count and percentage of hospitalizations that match between case mix and ResDAC
  - What percentage of demographic characteristics match?
  - Does the payer / dual eligibility status match across the data sources?

- Count and percentage of hospitalizations that do not match between case mix and ResDAC
  - What are the demographic and payer characteristics of those that don’t match?

- Explore data issues for subgroups with low match rates
Next Steps

• Start the implementation of the Phase 3 linkage process next month

• Use the linked database from Phase 3 to explore and produce more reliable and meaningful reporting on health system performance:
  • More accurate payer information
  • More accurate patient eligibility/enrollment information, including dual-eligibility status
  • Better payment data for various services
  • Improve/refine CHIA’s current reporting, such as annual reporting of hospital readmissions and revisits
  • More opportunity to produce additional and better reporting on health system performance and health care cost and trends
Questions?
For additional questions, please contact:

Zi Zhang, MD, MPH
Senior Director of Research
Center for Health Information and Analysis
zi.zhang@state.ma.us

Catherine Nwachukwu, MPH
Associate Manager of Research
Center for Health Information and Analysis
catherine.nwachukwu@state.ma.us