Use of Unique Beneficiary IDs in Medicaid Data Analyses

Medicaid Innovation Accelerator Program - National Webinar

January 25, 2018
3:00 PM – 4:00 PM EDT
Logistics for the Webinar

- All lines will be muted
- Use the chat box on your screen to ask a question or leave a comment
  - Note: chat box will not be seen in “full screen” mode
- Slides will be posted online
Welcome!

- Jessie Parker, GTL and Analyst on Medicaid IAP Data Analytic Team, Data and Systems Group, CMCS
Today’s Speakers

- Manjusha Gokhale, Senior Data Scientist, Truven Health Analytics, an IBM Company
- Bruce Greenstein, Chief Technology Officer, U.S. Department of Health and Human Services
Agenda for Today’s Webinar

- Introduction
- Overview of the Medicaid Innovation Accelerator Program (IAP)
- Working with Beneficiary Identifiers (IDs)
- Linkage Across Data Sources
- National Death Index
- Takeaways from Today’s Webinar
Medicaid Innovation Accelerator Program (IAP)

Medicaid Delivery System Reform

Program Areas

- Improving Care for Medicaid Beneficiaries with Complex Care Needs and High Costs
- Promoting Community Integration Through Long-Term Services and Supports
- Supporting Physical and Mental Health Integration
- Reducing Substance Use Disorders

Functional Areas

- Data Analytics
- Quality Measurement
- Performance Improvement
- Value-Based Payment and Financial Simulations
Goals for Today’s Webinar

In this interactive webinar, states will learn about:

- challenges in working with Medicaid enrollment data
- linkage methods
- linking to the National Death Index (NDI)
- examples of other linkages with state data
Use of Unique Beneficiary IDs in Medicaid Data Analyses

Challenges and Strategies

Manjusha Gokhale, Senior Data Scientist, Truven Health Analytics, an IBM Company
Beneficiary IDs in Medicaid Data

- Accurate identification of unique individuals is important for program administration, oversight, and analytics.

- Analyses which require correct denominator information include:
  - utilization analysis and comparison to benchmarks
  - assessment of expenditures
  - population health analysis
Medicaid Enrollee Identifier Assignment

- Medicaid enrollee identifiers are assigned by each state Medicaid agency.

- This identifier is assigned during enrollment along with highly identifiable information including:
  - social security number (SSN)
  - date of birth (DOB)
  - first name
  - last name
  - gender
  - address
Medicaid Enrollee Identifier Issues

- If you simply count the number of unique Medicaid enrollees identifiers in a year, you would likely get a number which was different than the total number of Medicaid enrollees.

- This is due to known issues with enrollment which include:
  - carve-outs for managed care, behavioral health, pharmacy coverage
  - combined mother/baby claims at birth
  - disenrollment / re-enrollment
Medicaid Enrollee Identifier Issues: Specialty Carve-outs

• Specialty carve-outs are arrangements where the state has contracted a third-party entity to administer the care given for certain services.

➢ Issue: Presence of multiple enrollee identifiers.

➢ Recommendation: Maintain a crosswalk of specialty carveout enrollee identifiers to state Medicaid enrollee identifiers.
Medicaid Enrollee Identifier Issues: Vertical Carve-outs

- **Vertical carve-outs** are where the state has contracted with an organization to administer care, such as Medicaid Managed care plans.

  - **Issue:** Individual is listed in Medicaid enrollment, but they could also be assigned another internal enrollment identifier by the health plan. The individual’s utilization is not in the Medicaid claims.

  - **Recommendation:** Maintain a crosswalk of vertical carveout enrollee identifiers to state Medicaid enrollee identifiers. Exclude these individuals in any cost or use analyses with Medicaid claims.
Medicaid Enrollee Identifier Issues: Combined Mother/Baby Enrollment

- **Mother/Baby**: Healthy babies are usually not enrolled in Medicaid at the time of birth.

  - **Issue**: Some current enrollment methods undercount healthy babies in Medicaid enrollment.

  - **Recommendation**: Confirm the number of infant enrollees by augmenting figures with information from birth records and hospital discharge claims.
Medicaid Enrollee Identifier Issues: Disenrollment/Re-enrollment

• Disenrollment/Re-enrollment: Some individuals will disenroll from Medicaid and later re-enroll and get assigned a different Medicaid enrollee ID.

  ➢ Issue: The same individual is represented several times in the enrollment data.

  ➢ Recommendation: Use Social Security Number to confirm that an individual does not have prior Medicaid enrollee ID.
Master Patient Index Definition

- **Master patient index** is a method of aggregating the information from disparate sources.

- The master patient index should contain only those fields which uniquely identify an individual (e.g. Medicaid ID, SSN, date of birth, gender).

- Ideally, Medicaid enrollee information should be consolidated into a master patient index.
Deterministic vs Probabilistic Matching

• Deterministic matches are exact matches
• Probabilistic matching uses a statistical approach and calculates the likelihood of a match as in the examples below:

- **Phonetics**
  - Katherine vs. Catherine

- **Synonyms**
  - Andrew = Andy
  - George = Jorge
  - 1st = First
  - Jr = Junior

- **Abbreviations**
  - Jerome David Salinger ~ J.D. Salinger

- **Concatenation**
  - Van de Velde = Vandevelde

- **Misalignment**
  - Robert John = John Robert

- **Edit Distance**
  - 876-5309 ~ 876-5390

- **Transliteration**
  - Toyota = トヨダ

- **Date Similarity**
  - 01/01/1973 ~ 01/02/1973

- **Proximity**
  - Geocodes and great-circle distance

- **Noise Words**
  - IBM Co. = IBM
Deterministic Matching

- Advantages
  - Confidence of match
  - Easy to understand and explain
  - Not dependent on knowledge of data file
  - Can use all matching fields and then drop criteria one-by-one to capture remaining non-matches

- Disadvantages
  - Rigid in structure
  - May undercount denominator
  - Can exclude common errors such as contractions of name, address changes
Probabilistic Matching

- Advantages
  - Can match across fields which may contain transcriptions, multiple spellings, address changes
  - Ability to maintain a longer longitudinal enrollment file

- Disadvantages
  - Difficult to describe
  - Can be hard / expensive to implement
  - May have false matches / non-matches
  - Highly dependent on patterns in database
# Master Patient Index

Sample Enrollment File

<table>
<thead>
<tr>
<th>ID</th>
<th>SSN</th>
<th>DOB</th>
<th>First Name</th>
<th>Last Name</th>
<th>Address</th>
<th>City</th>
<th>Date</th>
<th>Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>A948</td>
<td>123-45-6789</td>
<td>7/7/2007</td>
<td>Anita</td>
<td>Chen</td>
<td>5 Lexington Ave</td>
<td>Watertown, MA</td>
<td>1/1/17</td>
<td>Medical</td>
</tr>
<tr>
<td>M948</td>
<td>123-45-6789</td>
<td>7/7/2007</td>
<td>Anita</td>
<td>Chen</td>
<td>5 Lexington Ave</td>
<td>Watertown, MA</td>
<td>1/1/17</td>
<td>RX</td>
</tr>
<tr>
<td>A948</td>
<td>123-45-6789</td>
<td>7/7/2007</td>
<td>Anita</td>
<td>Chen</td>
<td>5 Lexington Ave</td>
<td>Watertown, MA</td>
<td>2/1/17</td>
<td>Medical</td>
</tr>
<tr>
<td>M948</td>
<td>123-45-6789</td>
<td>7/7/2007</td>
<td>Anita</td>
<td>Chen</td>
<td>5 Lexington Ave</td>
<td>Watertown, MA</td>
<td>2/1/17</td>
<td>RX</td>
</tr>
<tr>
<td>A948</td>
<td>123-45-6789</td>
<td>7/7/2007</td>
<td>Anita</td>
<td>Chen</td>
<td>5 Lexington Ave</td>
<td>Watertown, MA</td>
<td>3/1/17</td>
<td>Medical</td>
</tr>
<tr>
<td>M948</td>
<td>123-45-6789</td>
<td>7/7/2007</td>
<td>Anita</td>
<td>Chen</td>
<td>5 Lexington Ave</td>
<td>Watertown, MA</td>
<td>3/1/17</td>
<td>RX</td>
</tr>
<tr>
<td>A948</td>
<td>123-45-6789</td>
<td>7/7/2007</td>
<td>Anita</td>
<td>Hines</td>
<td>5 Lexington Ave</td>
<td>Watertown, MA</td>
<td>4/1/17</td>
<td>Medical</td>
</tr>
<tr>
<td>M948</td>
<td>123-45-6789</td>
<td>7/7/2007</td>
<td>Anita</td>
<td>Hines</td>
<td>5 Lexington Ave</td>
<td>Watertown, MA</td>
<td>4/1/17</td>
<td>RX</td>
</tr>
<tr>
<td>M948</td>
<td>123-45-6789</td>
<td>7/7/2007</td>
<td>Anita</td>
<td>Hines</td>
<td>6 Watertown Ave</td>
<td>Lexington, MA</td>
<td>5/1/17</td>
<td>Medical</td>
</tr>
<tr>
<td>A948</td>
<td>123-45-6789</td>
<td>7/7/2007</td>
<td>Anita</td>
<td>Hines</td>
<td>6 Watertown Ave</td>
<td>Lexington, MA</td>
<td>5/1/17</td>
<td>RX</td>
</tr>
<tr>
<td>M948</td>
<td>123-45-6789</td>
<td>7/7/2007</td>
<td>Anita</td>
<td>Heinz</td>
<td>6 Watertown Ave</td>
<td>Lexington, MA</td>
<td>5/1/17</td>
<td>Mgd Care</td>
</tr>
</tbody>
</table>
Deterministic vs Probabilistic Matching

cont.

- Deterministic matching on SSN would result in a single person – e.g., Anita Heinz.

- However, if we did not have SSN or DOB and matched on last name, first name, address and city, we would end up with three people – Anita Chen, Anita Hines, and Anita Heinz.

- Probabilistic linkage would allow us to have one person – Anita Heinz – even without SSN or DOB.
Polling Question

Has your state agency used any of the following when working with beneficiary IDs?

- Probabilistic matching
- Deterministic matching
- We have used both probabilistic and deterministic matching methods
- We have not used either approach
Recommendations

- Establish a hierarchy of linkage
- Examine the matches for confirmation and examine a set of non-matches to view patterns in errors
- Loosen the match criteria and check to see whether correct people matched
Recommendations (cont’d)

- Enrollment data which is prone to transcription errors and name changes would be a good candidate for probabilistic linkage.

- When creating a master patient index, think about the purpose of creating such a file (e.g., longitudinal analysis).

- Compare results to previously reported benchmarks.
Linkage to Other Sources

- Once the Master Patient Index is created, one can use this to link to a number of different sources including administrative health claims, electronic health records, vital statistics and others.

- Both deterministic and probabilistic techniques can be used to linkage between data sources.
Linkage to National Death Index (NDI)

- The CDC National Death Index is a nationwide compilation of state death records
- Researchers can apply to use the Index
- If approved, the researchers send the National Center of Health Statistics (NCHS) an password-protected encrypted file with identifying fields using the structure specified
- NCHS matches the state research file with the NDI and returns results files
Results from NDI Linkage

- Results from NDI Linkage will be complex with multiple files and multiple linkages to a single person.
- CDC provides guidance on how to interpret your results
- Reference: https://www.cdc.gov/nchs/ndi/index.htm
Example 1: Selecting NDI Data

1. **Core Database**
   - Divide Core Database into 3 Subsets:
     - **Unknown Vital Status**
     - **Known Living**
     - **Known Deceased**

2. **Unknown Vital Status**
   - All With Unknown Status
   - Send Combined File of 3 Subsets to SSA.
   - **SSA Results**
     - **Unknown Status Group**
       - SSA vital status code, e.g., living, deceased, unknown
       - Add SSA Vital Status Data to Core Database.
         - SSA Vital Status Confirmed Living
           - No Further Processing
         - SSA Vital Status Not Confirmed Living
           - Send to NDI.

3. **Known Living**
   - **Random Sample of Known Living**
     - Sample of Known Living
       - For SSA vital status living, calculate percentage of successfully matched living.
     - For SSA vital status of deceased or unknown, review discrepancies (considered “false alives”).

4. **Known Deceased**
   - **Random Sample of Known Deceased**
     - Sample of Known Deceased
       - For SSA vital status living, review discrepancies (considered “false alives”).
       - Any other SSA vital status is acceptable.
## Example 1: Assessing Match

<table>
<thead>
<tr>
<th>Strata</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exact match</td>
</tr>
<tr>
<td>2</td>
<td>Exact SSN and sex match</td>
</tr>
<tr>
<td>3</td>
<td>Exact SSN match</td>
</tr>
<tr>
<td>4</td>
<td>8-digit SSN and sex match</td>
</tr>
<tr>
<td>5</td>
<td>7-digit SSN and sex match</td>
</tr>
<tr>
<td>6</td>
<td>6-digit SSN and sex match</td>
</tr>
<tr>
<td>7</td>
<td>5-digit SSN and sex match</td>
</tr>
<tr>
<td>8</td>
<td>Valid user SSN, missing NDI SSN, name/DOB/sex match</td>
</tr>
<tr>
<td>9</td>
<td>Valid user SSN, missing NDI SSN, name, sex, DOB month and day match, DOB year within 1</td>
</tr>
<tr>
<td>10</td>
<td>Valid user SSN, missing NDI SSN, phonetic name, DOB, and sex match</td>
</tr>
<tr>
<td>11</td>
<td>Name, DOB, and sex match</td>
</tr>
<tr>
<td>12</td>
<td>Name, sex, DOB month and day match, DOB year within 1</td>
</tr>
<tr>
<td>13</td>
<td>Phonetic name, DOB, and sex match</td>
</tr>
<tr>
<td>14</td>
<td>Exact DOB match</td>
</tr>
<tr>
<td>15</td>
<td>Last name, first name, DOB month, sex match, DOB year within 10</td>
</tr>
</tbody>
</table>
### Example 1: Results

<table>
<thead>
<tr>
<th>1 (Strata)</th>
<th>2 (Criteria)</th>
<th>3 (No. of NDI Records Meeting Criteria)</th>
<th>4 (No. of Keeper Death Certificates Requested From States)</th>
<th>5 (No. of Death Certificates Received From States)</th>
<th>6 (No. of Death Certificates of Individuals Kept After Manual Screen)</th>
<th>7 (No. of Death Certificates of Individuals Accepted Into Study Cohort)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exact match</td>
<td>1,778</td>
<td>1,778</td>
<td>1,778</td>
<td>1,778</td>
<td>1,778</td>
</tr>
<tr>
<td>2</td>
<td>Exact SSN and sex match</td>
<td>1,173</td>
<td>1,130</td>
<td>1,103</td>
<td>1,103</td>
<td>1,035</td>
</tr>
<tr>
<td>3</td>
<td>Exact SSN match</td>
<td>159</td>
<td>116</td>
<td>99</td>
<td>74</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>8-digit SSN and sex match</td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>69</td>
</tr>
<tr>
<td>5</td>
<td>7-digit SSN and sex match</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>6-digit SSN and sex match</td>
<td>21</td>
<td>18</td>
<td>17</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>5-digit SSN and sex match</td>
<td>105</td>
<td>88</td>
<td>80</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>Valid user SSN, missing NDI SSN, name/DOB/sex match</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>
## Example 1: Results (continued)

<table>
<thead>
<tr>
<th>1 (Strata)</th>
<th>2 (Criteria)</th>
<th>3 (No. of NDI Records Meeting Criteria)</th>
<th>4 (No. of Keeper Death Certificates Requested From States)</th>
<th>5 (No. of Death Certificates Received From States)</th>
<th>6 (No. of Death Certificates of Individuals Kept After Manual Screen)</th>
<th>7 (No. of Death Certificates of Individuals Accepted Into Study Cohort)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Valid user SSN, missing NDI SSN, name, sex, DOB month and day match, DOB year within 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>Valid user SSN, missing NDI SSN, phonetic name, DOB, and sex match</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>Name, DOB, and sex match</td>
<td>45</td>
<td>39</td>
<td>37</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>12</td>
<td>Name, sex, DOB month and day match, DOB year within 1</td>
<td>43</td>
<td>31</td>
<td>29</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>Phonetic name, DOB, and sex match</td>
<td>80</td>
<td>70</td>
<td>56</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>14</td>
<td>Exact DOB match</td>
<td>167</td>
<td>123</td>
<td>105</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>Last name, first name, DOB month, sex match, DOB year within 10</td>
<td>8,396</td>
<td>1,060</td>
<td>956</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>12,094 records</td>
<td>4,580 records</td>
<td>4,386 death certificates</td>
<td>3,157 death certificates</td>
<td>3,033 death certificates</td>
<td></td>
</tr>
</tbody>
</table>
Example 2: Opioid Data Mapping in Massachusetts
## Example 2: Opioid Linkage

<table>
<thead>
<tr>
<th>Match Level</th>
<th>Identifiers To Be Matched</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exact match on first name, last name, Social Security number, gender, birth date, street address #1, street address #2, town of residence, and zip code.</td>
</tr>
<tr>
<td>2</td>
<td>Exact match on last name, Social Security number, gender, birth date, town of residence, and zip code.</td>
</tr>
<tr>
<td>3</td>
<td>Exact match on Social Security number, gender, and birth date.</td>
</tr>
<tr>
<td>4</td>
<td>Exact match on first name, last name, gender, birth date, street address #1, street address #2, town of residence, and zip code.</td>
</tr>
<tr>
<td>5</td>
<td>Exact match on first name, last name, gender, birth date, town of residence, and zip code.</td>
</tr>
<tr>
<td>6</td>
<td>Exact match on first name, last name, gender, and birth date.</td>
</tr>
<tr>
<td>7</td>
<td>Exact match on first name, last name, gender, and birth date.</td>
</tr>
<tr>
<td>8</td>
<td>First and third letters of first name, first and third letters of last name, gender, birth date</td>
</tr>
<tr>
<td>9</td>
<td>Street address #1, street address #2, town of residence and zip code</td>
</tr>
<tr>
<td>10</td>
<td>Exact match on first name, last name, and birth date</td>
</tr>
</tbody>
</table>
References for Examples

• Example 1:


• Example 2:

Commonwealth of Massachusetts Chapter 55 Opioid Report. Available at:

https://www.mass.gov/service-details/chapter-55-overdose-report
Bruce D. Greenstein,
Chief Technology Officer,
U.S. Department of Health and Human Services
@HHSCTO
About HHS

- The U.S. Department of Health and Human Services (HHS) is the nation's principal agency for protecting the health of all Americans and providing essential human services. The Department includes CMS, CDC, FDA, NIH, AHRQ, HRSA, SAMHSA

- ~ 79,540 employees

- The Office of the Chief Technology Officer, located in the Immediate Office of the Secretary, provides leadership and direction on data, technology, innovation and strategy across the Department of Health and Human Services

Source: hhs.gov
The Office of the Chief Technology Officer

**Data Insights Institute**
Leveraging data to generate better health and human services insights
- HHS ReImagine Initiative
- Code-a-thons
- Enterprise Data Strategy

**Industry Partnerships**
Driving partnerships across external, international, and HHS
- Reimagine RO
- Health Datapalooza
- Kidney Innovation Accelerator (ASN)
- Global Digital Health Partnership
- Start Up, Entrepreneurial, Private-Public Partnerships

**Innovation IDEA Lab**
Leadership, Council, and Execution
- HHS Ignite Accelerator
- Secretary’s Ventures Fund
- Entrepreneur-in-Residence
- HHS Open Innovation
Louisiana State Overview

- Population is approximately 4.6 million
- Strong Governor Model – Cabinet Agencies
- Department of Health includes Medicaid, Medicaid Eligibility, Public Health, Behavioral Health, Aging
- Doesn’t Include Human Services, Eligibility
- Constant struggle to keep up with budget pressures
State Priorities

• Reduce provider burden
• Minimize human capital requirements to state agencies
• Rising to challenges of new system procurements, and regulatory complexities
• Implementing new laws and programs
Programmatic Structure

• Medicaid Program – move to managed care
• Programs within the department vs. programs in other departments
• Programs within Medicaid
• Programs run by Public Health Department
• Programs in other departments
Program Data vs Enterprise Data

- Beyond Eligibility Data – A dualopoly
- How far does the Medicaid ID go?
- What is the common client index?
- Health Plans – where do they fit? How is the data handled?
- Same people, many programs
Practical Applications

- Death data, opioid reporting, and validity
- Birth data, claims data, and outcomes
- Immunizations, health plans – registries and claims
Death and Beyond

- Vital records and data
- Capturing and sharing death data
- Making vital record’s death module data more actionable
- Constructing a positive Return on Investment (ROI) for enhancing vital records systems
Try to Keep Up with Us

- @HHSCTO
- @HHSIDEALab
- www.hhs.gov/IDEALab
- IDEALab@hhs.gov
Questions?
Takeaways

• An awareness of administrative issues with Medicaid IDs can improve the success of an analytic project

• Probabilistic matching is often most appropriate for linking Medicaid IDs across administrative records

• Constructing a positive ROI for enhancing records systems may support evidence-based policy decisions
Thank You

Thank you for joining today’s webinar!

Please take a moment to complete the post-webinar survey.

For more information & resources, contact MedicaidIAP@cms.hhs.gov