PDMP Patient Matching Challenges and Opportunities

September, 2019

Vishal Patel, PhD
Director, Data Science R&D, Appriss
vpatel@appriss.com
@vishalrp @ApprissHealth
What Appriss Does Today

Founded in 1994, Appriss provides proprietary data and analytics solutions to address risk, fraud, safety and compliance issues for government and commercial enterprises worldwide.

**43 Statewide** programs delivering notification and information to crime victims

**Helping thousands of law enforcement to hold offenders accountable**

**Hundreds of state and federal agencies leverage Appriss data to make our nation safer and to prevent criminal fraud**

**51 PDMPs** depend on Appriss to deliver interstate information exchange around controlled substances (Opioids)

**43 PDMPs** have outsourced the management of their platforms to Appriss

Appriss provides the national platform (50 states) for preventing diversion of over the counter medicine containing pseudoephedrine

**More than 150,000 retail locations** use Appriss to mitigate fraud at the point of sale

**Many of the top retailers worldwide, across 35 countries, use Appriss to prevent loss and improve their bottom line**

**Appriss evaluates billions of transactions daily** as we prevent fraud and abuse within the retail world
Who are we?

• Appriss has been deploying linking for over 15 years across all its businesses
  • My background is in computer science, machine learning, graph algorithms, and bioinformatics
• Spent 10+ months on an R&D project to build the latest and greatest version of our patient matching engine—called ApprissID.

• Today, I’ll talk about various challenges around patient matching and how we solve them.
Why is patient matching important?

• Constructing a complete history of the patient’s prescription activity

• Affects risk scores and provider care
Terminology:

Linking
a.k.a. Consolidation
a.k.a. patient matching or entity resolution

Over-Linking
false positives or Type 1 error, matching or including records that don’t belong to the patient

Under-Linking
false negatives or Type 2, incomplete history or multiple patients provided when they are truly one patient
Patient Matching with an Example

- Address cleaning, zip code formatting, address standardization
- Name corrections: typos, nicknames, stay symbols, name derivations
- DEA & SSN number cleaning (number inversion, etc.)
- IDs linking/consolidation
Updating Patient Information

Patient Visit Record #6
Name: Sammy Lander
Address: 123 Main St
DOB: 4/30/1949
SSN: 552-14-9854
Pharmacy DEA: #DBD3488
Provider DEA: #AAP50220

- Provider determines patient visit #6 is also the same patient

ApprissID Created

ApprissID Updated

- All 6 records = 1 ID
PDMP at scale!

**PDMP Platform**

- **350 million** controlled substance prescriptions per year

**PDMP Interstate Sharing Hub**

- **262 million** transactions per month

**PDMP Clinical Workflow Integration**

- **82 million** patient encounters per month

NO COST to the states

43 PDMPs

51 PDMPs

40 PDMPs

Proprietary and Confidential © Appriss Health 2019
Challenges

Incorrect data entries
Typo
Bogus values
Lack of standards

Missing data
Phone < 65%
SSN < 5%

Use of diminutive names, maiden name

Address Changes

Error Observability Bias
Hardest Cases

• Twins
  • Same DOB, address, often visit the same provider/pharmacy
  • Similar first names and same last names

• Husband and Wife with the same DOB

• Father and Son with name suffix
Data Cleaning

• String Cleaning
• Name cleaning
• Address cleaning and normalization
• Standardize the formats of Phone Number, Zip etc
• Check for bogus numbers (entropy of numbers)
Different Industry Methods

• Deterministic Matching
• Probabilistic Matching
• Referential Data (using an MPI dataset for matching)
• Manual Matching

No single method works well, using a combination of methods produces high confidence matching.
How we combined it to create a robust solution?

• Deterministic Matching
• Probabilistic and Machine Learning Based Matching
• Referential Data
• Manual Matching
How do we measure errors?

Measuring the true error is almost impossible. We estimate the error based on the following indicators.

- QA Samples
- Support tickets
- Manual Consolidation
Issues and Continuous Collective Improvements

Support tickets

Every support ticket is researched and manual consolidations are analyzed to recommend improvements

Manual Consolidation

Improve Matching Algorithm of all PDMPs

Learnings from every support ticket are used to improve the algorithms in all states
Other Challenges

• Keeping data encrypted at all times.
• Ability to process a large number of records quickly.
• Real-time searching and update of information.
• High reliability and uptime.
• Horizontally scalable.
Conclusion

- Patient matching is a difficult problem- ApprissID solves it by creating a hybrid approach that combines multiple different techniques to achieve high confidence results.

- We have a framework in place designed for continuous improvement. Support tickets and manual consolidations are used to find improvements which are applied to all states.