



# PDMP Patient Matching Challenges and Opportunities

September, 2019

Vishal Patel, PhD

Director, Data Science R&D, Appriss

[vpatel@appriss.com](mailto:vpatel@appriss.com)

 [@vishalrp](https://twitter.com/vishalrp) [@ApprissHealth](https://twitter.com/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?

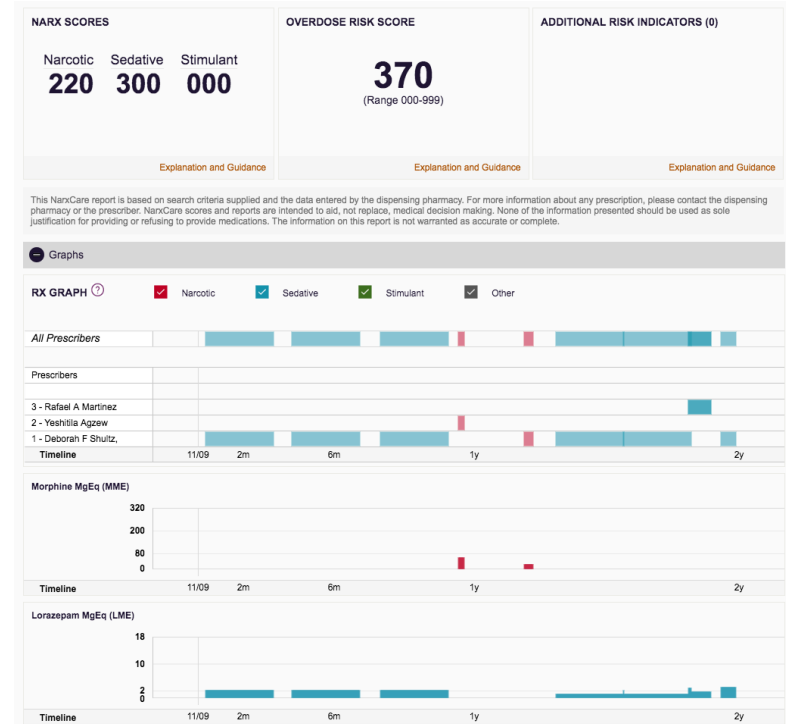


- Appriss has been deploying linking for over 15 years across all its businesses
  - My background is in computer science, machine learning, graph algorithms, and bioinformatics
- Spent 10+ months on an R&D project to build the latest and greatest version of our patient matching engine— called ApprissID.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



## 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

## Patient Visit Record #1

**Name:** Sammy Land  
**Address:** PO Box 454  
**DOB:** 4/30/1949  
**SSN:** 552-14-9854  
**Pharmacy DEA:** #DBD3488  
**Provider DEA:** #AAP50220

## Patient Visit Record #2

**Name:** Sammy Land  
**Address:** PO Box 455  
**DOB:** 4/30/1949  
**SSN:** 552-14-9854  
**Pharmacy DEA:** #DBD3487  
**Provider DEA:** #AAP50220

## Patient Visit Record #3

**Name:** Sammy Landd  
**Address:** PO Box 454  
**DOB:** 4/30/1959  
**SSN:** 552-14-9854  
**Pharmacy DEA:** #DBD3848  
**Provider DEA:** #AAP50221

## Patient Visit Record #4

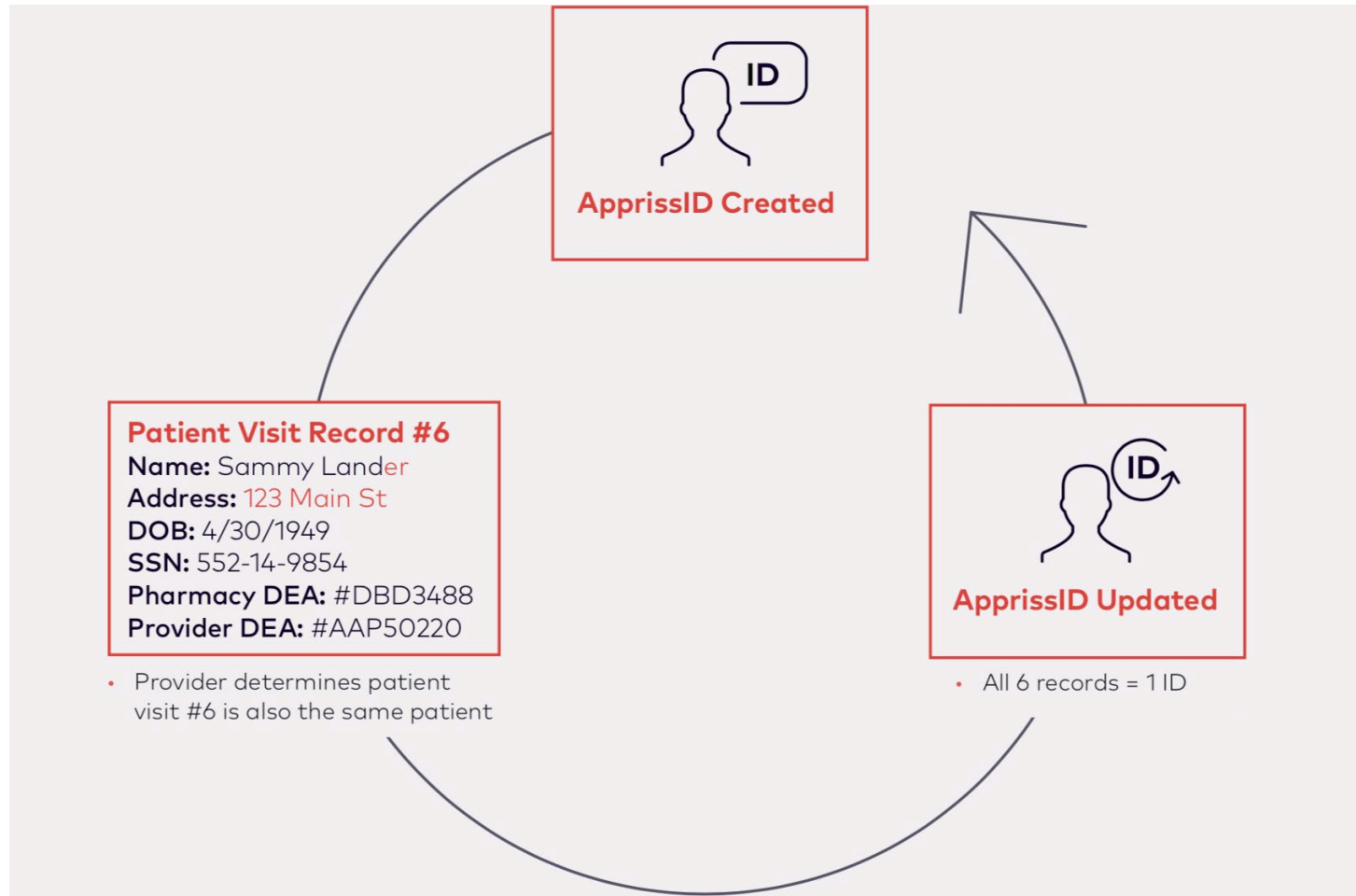
**Name:** Sammy Laannd  
**Address:** PO Box 454  
**DOB:** 4/30/1949  
**SSN:** 552-14-9854  
**Pharmacy DEA:** #DBD3488  
**Provider DEA:** #AAP502202

## Patient Visit Record #5

**Name:** Sammy Land  
**Address:** PO BOX 454  
**DOB:** 4/03/1949  
**SSN:** 552-14-9854  
**Pharmacy DEA:** #DBD34887  
**Provider DEA:** #AAP50220

- 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

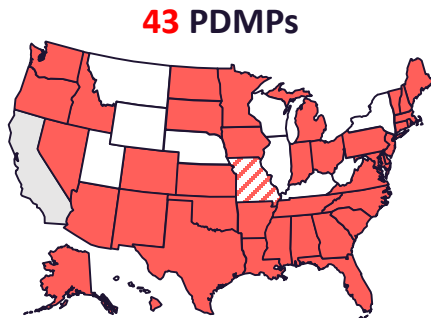
# Updating Patient Information



# PDMP at scale!

## PMP Awarxe™ PDMP Platform

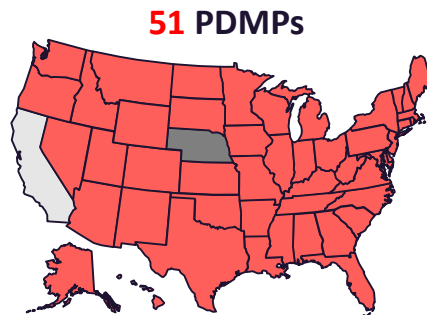
**350 million** controlled  
substance prescriptions  
per year



## PMP INTERCONNECT® PDMP Interstate Sharing Hub

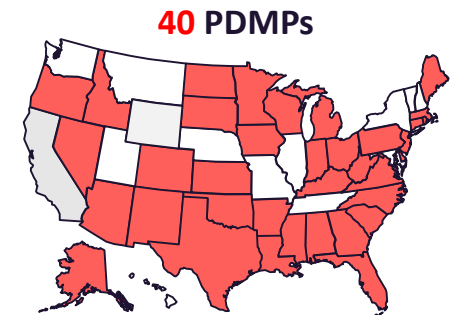
**262 million**  
transactions  
per month

**NO COST** to the states



## PMP Gateway™ PDMP Clinical Workflow Integration

**82 million** patient  
encounters per  
month





## 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

- 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

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

- 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

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

QA Samples



Support tickets

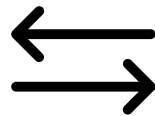


Manual Consolidation



# Issues and Continuous Collective Improvements

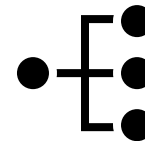
Support tickets



Manual Consolidation



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



Improve Matching Algorithm of all PDMPs

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

- 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.



- 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.