Resultant

Data-Driven Addiction Prevention and Recovery

The impact of the opioid crisis within the United States is well known and highly publicized. The cost—to individuals, families, and the economy—is unsustainable. Resultant's Data-Driven Addiction Prevention and Recovery (DDAPR) holistic methodology works to meet this complex public health crisis using a multi-pronged approach. DDAPR consultants analyze state, federal, and community data from macro and micro viewpoints to identify key levers for preventing citizens from becoming addicted to drugs and aiding in positive outcomes as they recover.

CONTENTS

97%

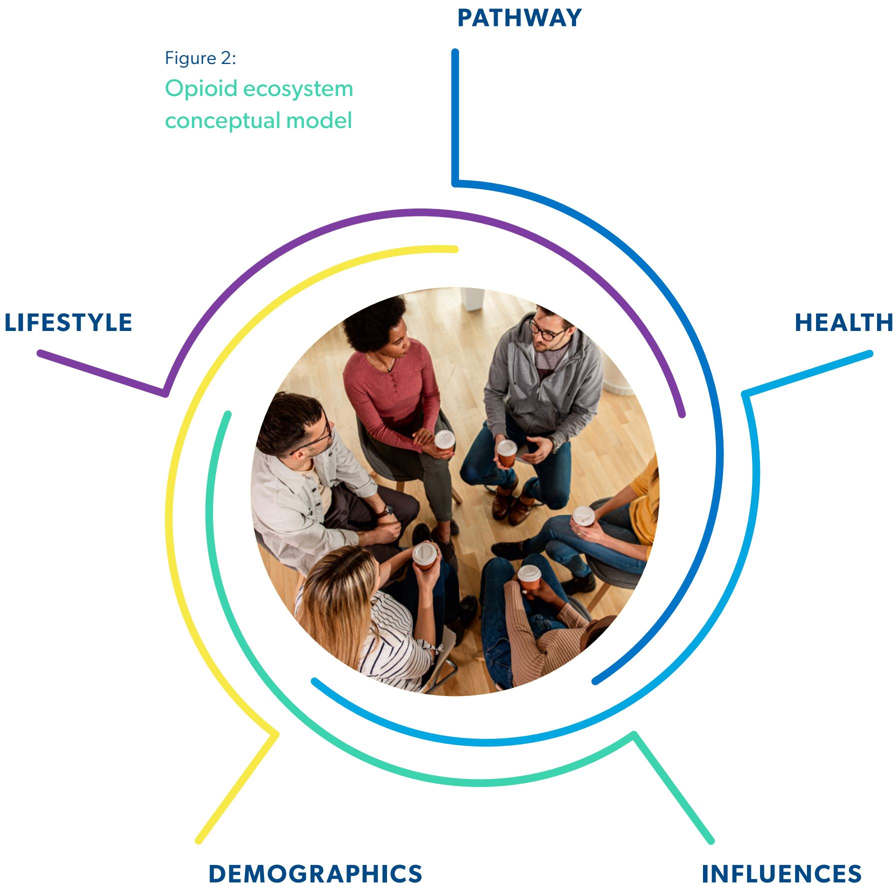
In practice, Resultant's algorithm has shown around 3.0% false negative rate and 0.6% false positive, yielding approximately 97% accuracy.

MACRO ANALYSIS

DDAPR's analytics methodology consists of parallel efforts intended to examine the opioid crisis from an individual and population level. The macro analysis workstream leverages technology such as graph databases to map and describe the comprehensive ecosystem of the opioid crisis. This work helps state agencies understand opioid use as it relates to psycho-ecological systems, medical diagnoses, and the mapping of a risk pathway to substance use. Agencies and programs can thus identify where intervention, prevention, and treatment are most effective for target populations.

Figure 1: DDAPR Methodology

Figure 2:
Opioid ecosystem conceptual model



MICRO ANALYSIS

The other side of DDAPR's analysis focuses on individuals, their friends, and their families. The micro analysis workstream empowers individuals by producing discrete, specific, and personalized information to citizens suffering from substance use disorder (SUD). DDAPR consultants produce predictive algorithms from state data and return risk-related information to citizens based on their individual characteristics, location, and behaviors. Tools developed from the micro analysis are tailored to empower citizens with unbiased, evidence-based recommendations.

KEY OUTCOMES

Leveraging the DDAPR methodology to combat the opioid crisis delivers these outcomes:

1. Holistic understanding of the SUD ecosystem in your citizen population
2. Identification of gaps in data collection, sourcing, sharing, and analysis
3. Improved collaboration between agencies not traditionally seen as players in the opioid crisis
4. Strategic efforts born out of evidence-based observations and predictions
5. More efficient service delivery to populations suffering from SUD
6. Separation of social and moral stigma from SUD prevention efforts

SIGNIFICANT OUTCOMES

- Identified optimal locations for new Opioid Treatment Programs; geocoded locations where patients overdosed, calculated drive time to the nearest existing or candidate locations, and used an optimization algorithm to pick the best locations
- Analyzed toxicology results for thousands of overdose deaths, determining how often different drugs were detected (in total and in combination) and whether the decedent had a recent legal prescription for that type of drug
- Created a heatmap showing where naloxone was given by first responders to patients who had overdosed on an opioid; developed a method to show geocoded locations of overdoses on the map while preserving privacy (see page 6 here)
- Performed a survival analysis to determine how individual and community variables affect outcomes (arrest, overdose, and death) after starting buprenorphine treatment
- Determined how often opioid-related deaths may not be classified correctly; created a machine learning model to classify deaths using text from the medical examiner and other data from death certificates
- Identified statistically significant increases in weekly overdoses by county
- Determined whether arrests are higher or not in counties that have an Opioid Treatment Program
- Evaluated whether overdoses increased or decreased after a syringe exchange program began in a county
- Performed analysis to predict whether an opioid overdose resulted in death within one year and which features were significant to the prediction.
- Developed analysis of significant factors for individuals who had been prescribed opioids during a given time period and then later experienced a fatal or non-fatal overdose

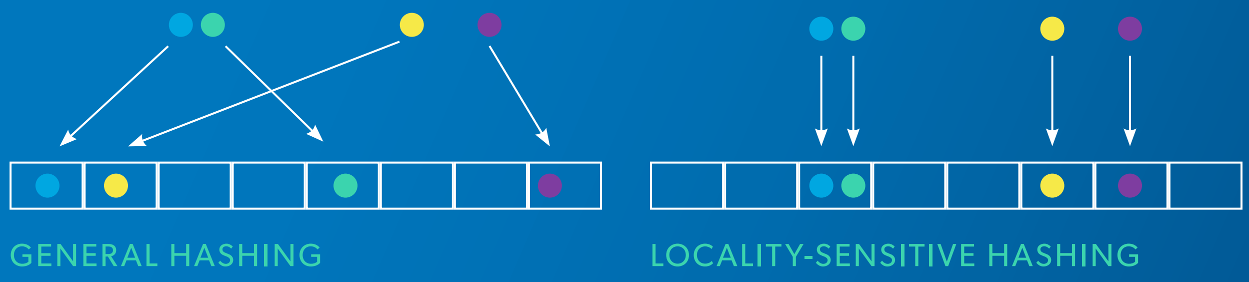


Figure 3: Matching patterns enable probabilistically similar hashes to be grouped

Record Linkage **as a Data Driver**

The key differentiator for the DDAPR project is the longitudinal database that captures data and events from multiple source systems. These disparate observations are then linked in order to generate a unified picture of an individual's actions and interactions with state programs. This allows the system to capture all relevant historical data and factors to SUD. To build such a longitudinal database for research purposes, Resultant employs a highly refined record linkage solution.

The objective of record linkage is to develop a comprehensive view of all relevant information pertaining to the same entity, whether a person, business, or event. Traditional record linkage methods leverage deterministic matching and require individuals have a unique and robust identifier, known as personally identifiable information (PII), such as their social security number or driver's license ID. If two records contain a common key, the records can be joined by exact matching. Where typos or mismatches exist, finding where records are linked becomes a manual process, meaning deterministic matching has distinct limitations.

Resultant has developed a unique methodology that combines a custom record linkage system with refined processes to probabilistically link large-scale datasets. While deterministic matching methods are powerful, they are also brittle. Source data must be clean, complete, and accurate in order to match records. Deterministic matching will capture only a percentage of possible matches. Resultant's probabilistic matching methods enable the state to get more value from data by evaluating the strength of a match on a spectrum. Records with a high likelihood of referencing the same individual can be successfully identified and matched where a deterministic match would fail.

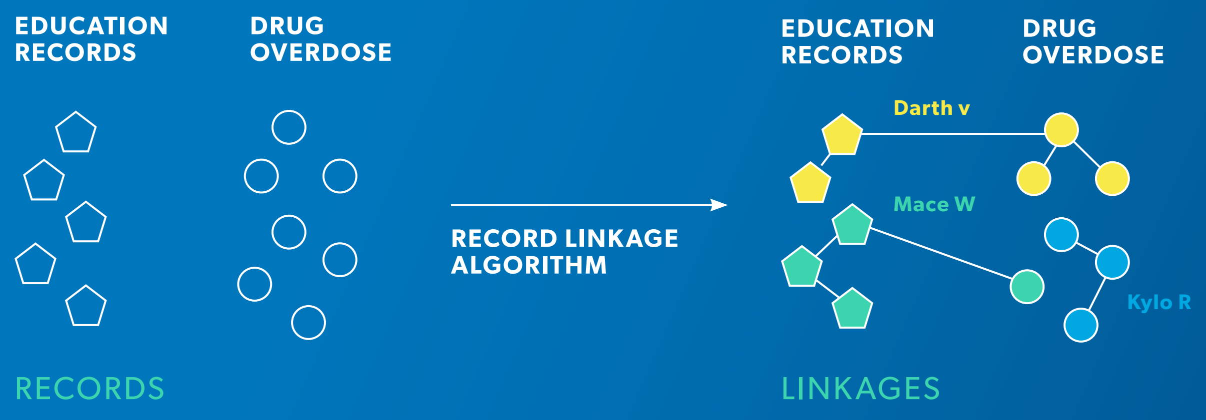


Figure 4: The algorithm identifies records across data sources

In the absence of a unique identifier, probabilistic record linkage connects PII and additional information to find matches. When missing keys, probabilistic record linkage holds distinct advantages over deterministic matching of PII that overcomes

- Data quality issues—typos, misspellings, missing or extra letters
- Data incompleteness—last four of social, year of birth, middle initial, nicknames
- PII mismatch, as when one database collects date of birth while another collects age
- Lifestyle changes in PII—marriage, change of address
- Source system requirement standard is variant between agencies (e.g. for one SSN is mandatory, other does not collect SSN or collects only last four digits of SSN)

Resultant's linkage solution focuses on using reliable record identifiers, such as PII and protected health information (PHI) values, and applying locality-sensitive hashing (LSH) onto the set of values; each identifier has a reliability score passed to the algorithm in the form of weights to the algorithm as a whole.

The solution is completely data-source and identifier agnostic. It can be configured to take *n*-amount of data sources and match with *n*-number of any reliable field. Given the configurable weight of a columnar identifier, the algorithm will train matching patterns from which the LSH will be applied. Thus, creating a linkage pool within which records with probabilistically similar hashes will be placed in a shared pool.

Finally, converting the data to a graph frame, we can assign a global ID to each graph node such that the same record can be identified across all data sources (including duplication within a data source).

In practice, Resultant's algorithm has shown around 3.0% false negative rate and 0.6% false positive, yielding approximately 97% accuracy.

Post- and Pre-Linkage **Validation**

Data cleansing and standardization before data goes into the algorithm improve the solution's effectiveness. Resultant has found great success by applying contextual business rules as the means to cleanse and standardize columnar values.

After the linkage algorithm successfully identifies global IDs for records, we recommend aligning the metadata surrounding those records. We've found great success in applying contextual business rules as the means to creating a single record view for all records sharing the same global ID, where each record stores the same value. When, for example, the global ID #1 has two addresses on two different records, the algorithm isn't stymied and will choose the appropriate record to display based on the preconfigured business rules.

Record linkage is the gasoline that drives the engine of data-driven approaches to solving public health issues. Government agencies, funding mechanisms, and programs organize around common problems and themes; unfortunately, society's most complex problems defy compartmentalization. Understanding the entire universe of characteristics that drive unwanted behavior and outcomes is critical. Resultant's record linkage solutions bridge the technical siloes that naturally occur, and give analysts dependable, robust datasets to find answers to our most serious problems.

ABOUT RESULTANT

Our team believes solutions are more valuable, transformative, and meaningful when reached together. Through outcomes built on solutions rooted in data analytics, technology, and digital transformation, Resultant serves as a true partner by solving problems with our clients, rather than for them.

DATA ANALYTICS

We help organizations understand their data landscape and solve problems by turning data into insight. While data can be dense, our team's empathetic approach to problem solving creates meaningful solutions with deep technical foundations.