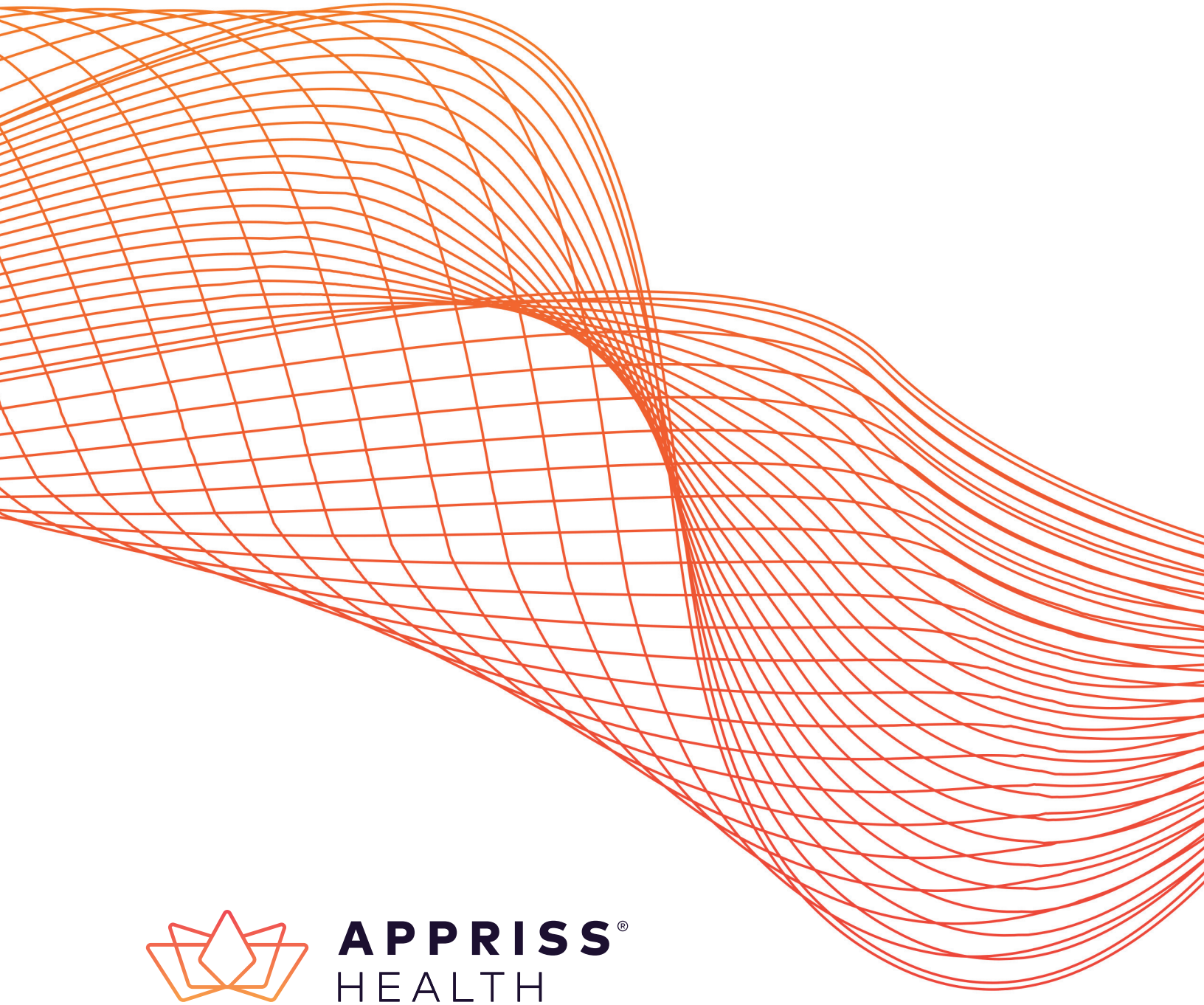


APPRISS® HEALTH

Addressing the Challenge of Accurate Patient-Matching for Prescription Drug Monitoring Programs

PATIENT CONSOLIDATION OVERVIEW

Linking medical records and creating a comprehensive view of a patient's medical history is vital to appropriate medical care. It is estimated that as many as one in five patients have incomplete or incorrect medical histories due to poor record linking.[1] Medical record linking encompasses lab test results, care notes, prescription fills, and more. Prescription drug monitoring programs (PDMPs) are a relatively small subset of the medical history that requires accurate record linking. PDMPs in the United States are maintained on a state-by-state basis and consist of all controlled substance prescriptions in addition to state-specified medications of interest. In addition to managing 43 state PDMPs, Appriss Health maintains an inter-state PDMP communication platform (PMP InterConnect) that allows for a patient's controlled-substance records across multiple PDMPs to be viewed by providers and pharmacists.

1. *https://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND_MG753.pdf*

43 managed state PDMPs

360 million PDMP searches conducted by **1.4 million** users each year

With 360 million PDMP searches conducted by 1.4 million users each year, ensuring that every record for an individual is returned without inadvertently grabbing information about another patient is a high-stakes balancing act that has been incredibly successful. Appriss Health's customer support team now fields about one call for every 770,000 PDMP patient searches reporting some issue or perceived issue with patient linking (0.00013% call rate). **Tailoring the patient linking algorithm to the reliability of the data entered into the PDMP allows for the best possible patient-matching.**

THE APPRISS HEALTH PATIENT-MATCHING ALGORITHM

Most patient record-linking approaches can be described as deterministic, probabilistic, referential, or a blend of all three like Appriss Health patient-matching. Taking the pieces of information in the patient record that identifies that individual a deterministic matching approach looks for exact matches between multiple records. For example, if the patient's name, date of birth, and Social Security number exactly match on two separate prescription records, those two records can be said to be for the same patient. A probabilistic approach to patient linking introduces a measure of uncertainty to the linking. This can be as simple as assuming that a patient named *"Staci"* and *"Stacey"* with

the same last name and date of birth are the same person, or more complicated and potentially risky, like linking two names based on how rare they are for the area. Appriss Health currently does probabilistic matching to attempt to link records where typos or name variants prohibit exact matching. If a record has the same name, date of birth and almost an exact match on a patient's phone number except for one digit, probabilistic matching would say that the probability of there being a typo present in the phone number is higher than the probability that there are two different people with the same name and date of birth, but that also have a phone number that differs by only one digit. Referential approaches to matching rely on external data sets that maintain lists of individuals or households, such as change of addresses databases, to be able to link records together. As more types of records from different sources are linked the likelihood of connecting patient histories from different regions of the country or different types of medical records does go up, but this can also increase the chances of linking two patients inappropriately, as field reliability can vary among data sets. Appriss Health patient-matching utilizes all the above techniques of matching in a manner which balances the riskiness of a mismatch with ensuring we capture all records belonging to an individual.

Table 1: Patient Linking Companies – algorithm types and source data

Product	Appriss Health patient-matching	Verato Universal MPI	Enterprise Master Patient Index	LeapMDM	Senzing Entity Resolution
Description	Referential database Probabilistic & deterministic matching	Referential databases Probabilistic & deterministic matching	Probabilistic & deterministic matching	Referential databases Probabilistic & deterministic matching	Principle-based entity resolution, data set does not require training
Sources Used	Referential database Probabilistic & deterministic matching	Referential database Probabilistic & deterministic matching	Referential database Probabilistic & deterministic matching	Referential database Probabilistic & deterministic matching	Referential database Probabilistic & deterministic matching

Edges are combinations of reliable personal identifiers that can connect two records, sometimes called N-tuples

Patient linking in the PDMP starts with data filtering and cleaning. Each prescription record contains multiple fields of information that can be used to uniquely identify a person such as name, date of birth, home address, etc. To confidently use these fields to link separate prescriptions together, data quality is first assessed. Sometimes address fields will be used to write notes on the patient, such as "Check ID" or a phone number will be entered as "999-999-9999". Filtering out these nonsense data entries is vital to avoiding incorrect patient linking. Once this filtering is done, additional data cleaning is still needed. Standardizing names to upper case, removing extra spaces, applying address standardization, and grouping addresses such that the algorithm can identify functionally identical information is done for appropriate record linking.

After the patient record has been filtered and cleaned, multiple high-confidence N-tuples, or edges, are created. **A single record can generate a unique string combination of first name, last name, date of birth, and first three digits of the home address zip code (FN-LN-DOB-ZIP3), but also another edge of first name, last name, date of birth, and Social Security number (FN-LN-DOB-SSN).** Within the PDMP records, the fields used to link patients include name, date of birth, home address, phone number, Social Security number, and the DEA numbers of the prescriber who wrote the prescription and the pharmacy who filled the prescription. **Only fields that have passed the reliability filtering contribute to the edge creation, so a single record can generate more than 10 different edges.**

Once all the patient records have contributed their edges, an identity graph approach is used to link separate records together. If record 1 has the same edge as record number 2, and record number 2 has an edge mirrored in record 3, then record 1 and 3 are also connected. In the example demonstrated in Figure 1, four different sets of patient identifiers are linked together using various edges. The name *"John"* and *"Johnny"* are connected by assessing the last name, date of birth, and phone number. This also connects two different home addresses. The home address provides sufficient evidence, in combination with the last name and date of birth, to connect the additional two sets of patient identifiers, or nodes. The more edges that connect multiple records, the

redundancy and confidence there is that the patient has been linked appropriately, so the cycle seen in the lower right helps confirm that *"John"* also goes by *"Johnny"*.

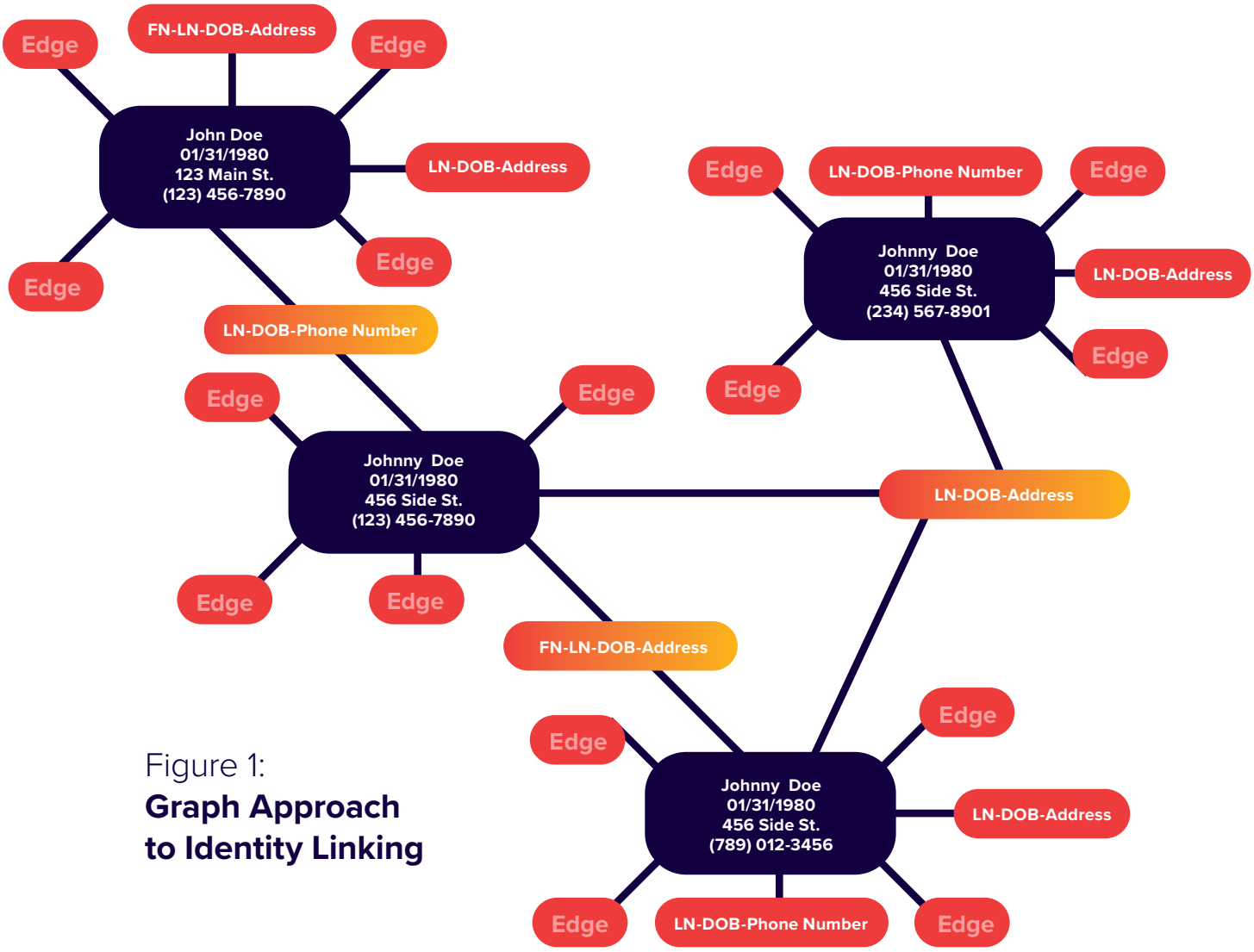


Figure 1:
Graph Approach to Identity Linking

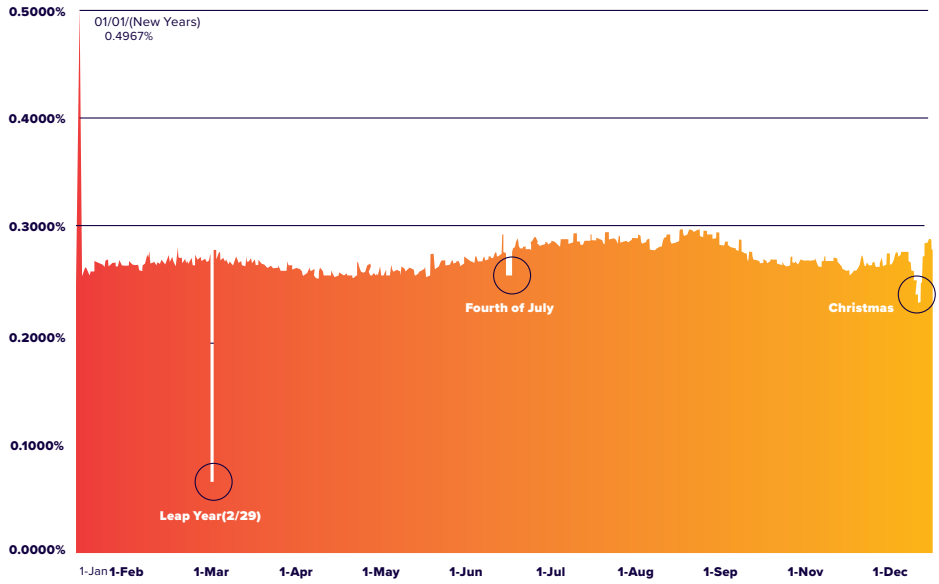
VALIDATING APPRISS HEALTH PATIENT-MATCHING

Assessing linking accuracy is a difficult metric to measure. A true gold-standard method of assessing patient record linking would be to go over the entire linked record set with each individual to confirm its accuracy. Lacking the ability to do a massive manual review with each patient, multiple methods are used to validate the patient linking. **Appriss Health receives about 2.9 billion searches a year for PDMP prescription histories across all the PDMPs it manages, and the customer support call center receives approximately 1 call about a patient linking issue for every 770,000 searches (0.00013%).** Another way to estimate error rate is to look for how many patient groups have ever had a manual edit to their linking structure. Altogether, **Appriss Health's linking has 176 million patient record groups and of these,** manual alterations to the linking (either decoupling

records or connecting separate groups) occur in about 1 out of every 2,997 groups (0.03%). The remaining 99.97% of patient groups have not had a manual change to the linking.

The patient-linking **issues that are most visible to end users are when two patients are accidentally linked together and yet have different birthdates.** Users expect that a person can have nicknames, name changes, and address changes but don't anticipate a patient having more than one birthdate. Yet data entry **errors, which are estimated to be responsible for up to 32% of incorrect patient linking, occur even in birthdates.**[2] *Figure 2* shows how often a PDMP prescription record has a given month/day birthdate combination. While there are some noticeable dips during leap years and major American holidays when birth rates are lower, there are also some notable peaks. "01/01" is almost twice as common as would be expected given the surrounding dates, and there are smaller peaks on the first of each calendar month as well, meaning patient birthdates are not always correctly entered into the PDMP. Appropriate patient linking within the PDMP can therefore include varying birthdates.

Figure 2: **Distribution of PDMP Prescription Birth Dates (Month/Day)**



Identifying which fields are linking together patient records that otherwise would not have been linked helps focus filtering and data cleaning efforts. If using the last name, date of birth, and zip code links

2. *https://pages.imprivata.com/rs/imprivata/images/Ponemon-Report_121416.pdf*

together the same records for the same patients as using first name, last name, date of birth and social security number, only including the edge that is most reliably and accurately entered helps avoid overlinking when inaccurate data is present. We tested the robustness of patient linking by systematically removing one type of edge from the patient-matching algorithm and re-linking *(Table 2)*. **Failing to use edges related to patient phone number had the largest impact on patient under-linking, with a 2.54% increase in the number of patient groups compared to a linking with all edges available.** Interestingly, not creating edges using data from a third-party change of address database only slightly (0.27%) increased the number of patient groups, as other combinations of patient identifiers generally captured the same information as the third party reference information.

Edge Description	Number of Patient Groups After Removing Edge	Increase in Number of Patient Groups
Last Name + DOB + Phone Number	3,266,203	102.54%
First Name + Last Name + DOB + ZIP3	3,232,603	101.49%
First Name + Last Name + DOB + Prescriber DEA	3,229,770	101.40%
First Name + Last Name + DOB + Phone Number Area Code	3,224,546	101.23%
First Name + Last Name + DOB + Pharmacy DEA	3,196,915	100.37%
Third Party Information + DOB	3,193,911	100.27%
First Name + Last Name + MMDD of DOB + ZIP5	3,187,979	100.09%
First Name + YYYY of DOB + SSN	3,185,882	100.02%
DOB + SSN	3,185,734	100.01%
All Edges Combined	3,185,270	

Table 2: A Selection of Edges Examined in the Edge Analysis for Appriss Health patient-matching Patient Linking on PDMP Records (Sample of N=52,241,926 dispensations)

Appriss' PDMP-tailored patient-matching algorithm **decreases the total number of patient groups by about 13%** compared to the third party's system.

Recently, one of the nationally recognized large-scale patient linking companies offered to do a head-to-head comparison of their linking capabilities compared to Appriss Health's internal PDMP linking algorithm. Using a sample set of about half a million patient records with multiple prescriptions across multiple states, linking was done via both algorithms. As shown in *Figure 3*, **1.1% of the third party's patient groups were under-linked in Appriss Health's linking scheme,** likely due to the third-party having access to a larger referential data set that allowed for linking two addresses or names. **2.5% of Appriss Health's patient groups were under-linked by the third-party scheme,** mostly due to Appriss Health being able to tailor our linking algorithm to PDMP data, particularly utilizing the frequently updated phone number field to generate high-confidence patient-matching. **Most patient groups (84.8%) were equivalent across patient linking algorithms.** Hand-reviewing cases where the two algorithms differed found no obvious cases of record overlinking on either algorithm's part. **Overall, Appriss Health's PDMP-tailored patient-matching algorithm decreases the total number of patient groups by about 13% compared to the third party's system.**

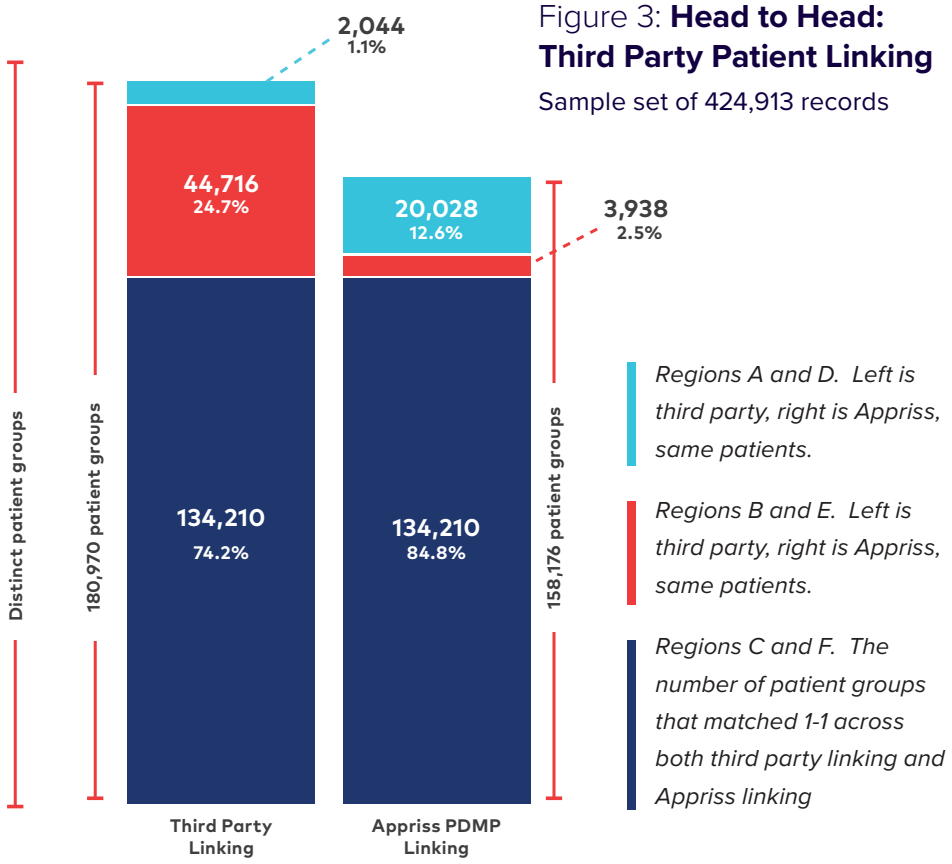


Figure 3: **Head to Head: Third Party Patient Linking**
Sample set of 424,913 records

Internally, the data science team dedicated to record linking regularly assesses and updates the field filtering, data cleaning, and edge creation code to improve patient linking. **As issues are found by the users and escalated to Appriss Health's customer support call center, a regular assessment of the reason why the issue occurred is logged and, if needed, the algorithm Is edited to minimize these errors.** One such algorithm change occurred early in the design process, when date of birth could vary by one character to try to link together records where there were data entry typos. The customer support then fielded a handful of calls where a father and son were connected via our patient linking algorithm. After investigation, we figured out that in the US, it's not uncommon for a son who was born on his father's birthday to also be given his name. Now knowing this naming practice, the Appriss Health patient-matching data science team altered the linking algorithm so that it only looked for typos in the final digit of the birth year, not decade, and the linking issues were resolved.

Twins are also difficult for accurate patient linking. Born on the same birthdate and possibly sharing a home address, and phone number, there are many identifying fields that exactly match between twins. Even more difficult for patient linking algorithms is that many twins are often named in rhyming or matching first names. For example, *"Jayden"* and *"Kayden"*, *"Madison"* and *"Mason"*, or *"Taylor"* and *"Tyler"*, in conjunction with all the other matching identifier fields, all look to a computer like they could have been data entry typos for a single patient. Continuous tweaks to the patient-matching algorithm have been made to avoid overlinked twins while still capturing name variations and typos that would underlink other patients.

Appriss Health's continually monitors patient linking, looking for ways to tailor the filtering, data cleaning, and linking process to improve PDMP linking accuracy. **If there is an edge that is consistently overlinking patients the algorithm is adjusted to reduce errors. Additional fields are also assessed for potential patient linking gains.** Middle name is an available field within PDMP data that could help link maiden and married names for some women but can also be unreliable, with pharmacy notes for the patient sometimes entered instead of a name. Some states are also now recording the driver's license of the

person picking up a prescription, which also could also help link together underlinked patient histories. Yet even a national patient ID would not completely solve the need for patient linking in medical records. Human error in data entry is always an issue, and any patient records that occur before such a national patient ID was put in place would still need to be linked.

CONCLUSION AND FUTURE DIRECTIONS

Patient-matching is a difficult technical problem to solve. As more records are linked from disparate sources, a more comprehensive patient history can be captured but at the same time, more errors can be introduced. **Appriss Health does PDMP linking both within and between more than 40 states and territories and have been doing so for more than 4 years. Over 12 million PDMP search requests are processed each day on more than 2.6 billion prescription records. Our best assessments of patient linking accuracy suggest that this linking is very accurate, with 0.00013% of searches generating a support call and 0.03% of patient groups overall having ever had manual intervention to fix linking.** Appriss Health's intimate knowledge of PDMP data allows for a tailored patient linking algorithm that creates for the best possible patient-matching.