IDENTIFICATION OF PROBLEM OPIOID USE RISK FROM ELECTRONIC HEALTH RECORDS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

MSc.

by

Jeffrey Hao-Chan Huang, Xinyuan Liu

May 2020

ABSTRACT

A study cohort of 2733 chronic opioid therapy patients was extracted from the electronic health record system (EHR) of Weill Cornell Medicine over the period of 2000 to 2018. The case group of 422 patients in this cohort who had developed problem opioid use (POU), including opioid dependence, opioid misuse, and opioid abuse, was subsequently identified by parsing recorded ICD-9 and ICD-10 codes. We proceeded to extract 31420 potential risk factors from this structured EHR data then encoded them as a binary feature vector for each patient. Pearson's Chi-square test was conducted on each potential risk factor between the case and control groups to identify a subset of 2860 potential risk factors with the highest probability of being correlated with an increased risk of developing POU. A logistic regression predicting the development of POU performed on this subset of risk factors achieved an area under the receiver operating characteristic curve (AUC) of 0.796. We then applied Recursive Feature Elimination to further reduce this set of risk factors to an optimal subset of 1150 features. A logistic regression performed on this optimal set of features achieved an AUC of 0.793. The features with the greatest positive coefficients in this regression model were mapped back to their respective concept domains specified by the Observational Medical Outcomes Partnership (OMOP) standardized vocabularies. The distribution of features among the concept domains indicates that the medical conditions suggested by a patient's symptoms, the drugs prescribed to a patient, and the medical procedures ordered for a patient captured by the EHR can be leveraged to detect an increased risk of developing POU.

CHAPTER 1

INTRODUCTION

In the late 1990s, a paradigm shift occurred which encouraged physicians to proactively address and treat pain [1, 3, 7, 8]. As a result of efforts to adequately manage this symptom, there has been a significant increase in the frequency at which physicians prescribe opioids, with approximately 1.9 million new initiates per year [3].

This increase in access to prescription opioids has been associated with a rapid escalation in the number of patients developing opioid dependence and problem opioid use (POU), defined as any use of opioids outside of prescription parameters [3]. Since regular opioid use results in dependence, patient populations receiving chronic opioid therapy (COT) have shown rates of POU as high as 40% [3, 5]. In response, an array of screening tools have been developed to enable clinicians to assess the risk of future problem opioid use for patients undergoing chronic opioid therapy (COT) [2, 3, 5, 8].

Although many of these tools have been validated, the time and resources required for a clinician to administer these tools render them impractical in clinical settings, such as primary care clinics, with time constraints and competing demands [4, 5]. Moreover, the transition from short-term opioid therapy to COT is often difficult to pinpoint. As a result, screening for the risk of future problem opioid use has not been reliably carried out [4].

However, research has identified several factors predictive of increased risk of POU such as age, sex, smoking status, opioid therapy type and dosage, prior mental health disorder diagnoses, prior documented opioid misuse, and medical comorbidities [4, 5, 9]. In medical systems that have implemented an electronic health record (EHR) system, these factors are inputted into then updated

in the EHR at each patient encounter [4, 5, 6, 9, 10].

Despite the availability of longitudinal health data in EHR systems and recent advances in machine learning techniques for drawing insights from large datasets, it remains unclear how these data sources should be leveraged to identify increased risk of developing POU from patient characteristics.

Therefore, we aim to aggregate longitudinal clinical data of patients undergoing COT from 2000 to 2018 at Weill Cornell Medicine, a tertiary referral medical center located in New York City. Since individual patient data stored in the EHR systems at Weill Cornell Medicine had been generated through the patient's clinical encounters, it may include records of medical history, observed signs and symptoms, prescribed medications, as well as ordered medical laboratory tests and results. The subset of COT patients who developed POU over this time period are also identifiable from this EHR data.

Once this data has been consolidated, we proceed to explore whether a binary encoding scheme is sufficient in generating a feature vector for each patient undergoing COT that can effectively capture the risk factors associated with increased risk of developing POU. We will assess whether the features generated through this process are predictors of increased risk for POU by analyzing the performance of logistic regression classifiers in identifying POU patients from the cohort of patients undergoing COT.

A note about this report: it is the final report for the Specialization Project, required for the Master of Science program in Information Systems with a concentration in Health Tech. This project was a two-person research project conducted under the advice of a faculty member at Weill Cornell Medical College.

CHAPTER 2

RELATED WORK

A longitudinal study was conducted by Hylan et al.[4] among 2752 chronic noncancer pain patients initiating COT, defined as the prescription of a minimum of 70-days' supply of a transdermal or oral opioid in a given calendar quarter, between 2008 and 2010 then tracked through the end of 2012. The participants of this study were patients enrolled in a large mixed-model health plan in Washington state which uses the Epic EHR system to document primary, specialist, and emergency care encounters as well as hospital discharge summaries. This study demonstrated that for a set of baseline risk predictors extracted from the EHR system, the odds ratio obtained by running logistic regression, with the dependent variable set as the presence of subsequent problem opioid use, can be used as a measure of the increase in risk of future problem opioid use given a positive predictor value. The authors utilized these odds ratios to derive a weighted count of present risk factors, which include: age group, smoking status, prior mental disorder diagnosis, prior alcohol dependence diagnosis, prior opioid dependence diagnosis, as well as prior hepatitis C diagnosis. They then proceeded to construct a simple predictive model, by setting an appropriate decision threshold. The predictive model derived from the weighted count of risk indicators resulted in an AUC value of 0.733 (C.I. = 0.666, 0.800) for the training set, 0.718 (C.I. = 0.666, 0.800) for the testing set, and 0.724 (C.I. = 0.684, 0.764) for the entire sample.

The Veterans Health Administration's Stratification Tool for Opioid Risk Mitigation [2] with AUC > 0.8 further demonstrates that similar predictive models trained on longitudinal patient EHR data can be applied on a daily basis to updated EHR entries, thereby automating the assessment of patient risk for fu-

ture problem opioid use. This system was deployed to assist VA providers with opioid risk evaluation and mitigation for patients. However, although VHA is a diverse and national system of 153 hospitals and more than 750 outpatient clinics, the system serves the sickest and poorest among the twenty-three million US veterans, many of whom have multiple chronic conditions. Thus, developing comparative effectiveness to the general population from this tool remains a challenge [11].

In a proof-of-concept study conducted by Reps et al. [12], the researchers first proposed then evaluated a standardized framework for patient-level prediction, implemented using the Observational Medical Outcomes Partnership (OMOP) Common Data Model and its standardized vocabularies. They followed their proposed framework to extract variables from the dataset which were subsequently used to develop a range of predictive models. The researchers then trained an L1-regularized logistic regressor using 3-fold cross-validation for each of 21 different outcomes within a target population of people suffering from depression across 4 observational databases. This approach led to highly-discriminative models for the outcomes such as ventricular arrhythmia and sudden cardiac death, as well as hypothyroidism across the 4 databases, with AUCs ranging between 0.732–0.808 and 0.763–0.845, respectively. However, for the outcomes of diarrhea and tinnitus, the discrimination was consistently poor across the dataset, ranging between 0.636–0.682 and 0.576–0.696, respectively. In general, these results suggest that observational databases can be utilized to develop clinically useful models that predict some outcomes, but certain types of outcomes may need more advanced methods or alternative datasets.

CHAPTER 3

METHOD

1. Identification of the Study Cohort

Our study cohort consists of 2733 patients identified as undergoing COT, extracted from the EHR of Weill Cornell Medicine among approximately three million patients over the period of 2000 to 2018. Every patient in the cohort fulfills the following definition of COT: at least 3 successive opioid prescriptions over a minimum time period of 3 months, with at least 16 tablets per prescription. We employed this definition of COT because it approximates the recommendation provided by the 2015 Agency Medical Directors' Group (AMDG) Interagency Guideline on Prescribing Opioids for Pain.

2. Identification of POU Patients Using ICD Codes

Since the term Problem Opioid Use (POU) has been used to describe a spectrum of clinically-observable behaviors and disorders, we defined POU to include opioid dependence, misuse, as well as abuse. From this definition, we used the corresponding ICD-9 and ICD-10 codes shown in Table 1 to extract a case group consisting of 422 POU patients from the study cohort using SQL queries.

ICD-9 Codes	Condition
305.5	Non-dependent opioid abuse
304.0	Opioid type dependence
304.7	Combination of opioid type drug with any other drug dependence
ICD-10 Codes	**Condition**
F11.1	Opioid abuse
F11.2	Opioid dependence
F11.9	Opioid use, unspecified

Table 1: ICD-9 and ICD-10 codes used to identify POU patients from the EHR.

3. EHR Data Transformed to the OMOP Common Data Model

Since data from clinical encounters may be stored across separate databases, we conducted our study on Weill Cornell Medicine's EHR data which had first been transformed into the OMOP Common Data Model. The OMOP Common Data Model consolidates data from disparate sources into a common format, in which clinical and other healthcare-related concepts are represented as a set of standardized vocabularies. Each distinct concept is encoded as a unique concept id [13].

To aggregate the EHR data collected for our study cohort, we processed the records associated with each patient within 6 selected standardized clinical data tables: *Visit_Occurrence*, *Observation*, *Condition_Occurrence*, *Procedure_Occurrence*, *Drug_Exposure*, and *Measurements*, which contain the core information gathered from longitudinal clinical events [13]. These records include the setting and time span of each clinical encounter, the patient's symptoms and the medical conditions they may suggest, the medical procedures either ordered or carried out, the medications prescribed, as well as the results of laboratory tests. Each record in each of these standardized clinical data tables is associated with a patient recorded in the *Person* table by the foreign key *person_id*.

The entity-relationship diagram specifying the relationships between the standardized clinical data tables of the OMOP Common Data Model is shown in Figure 2 below.

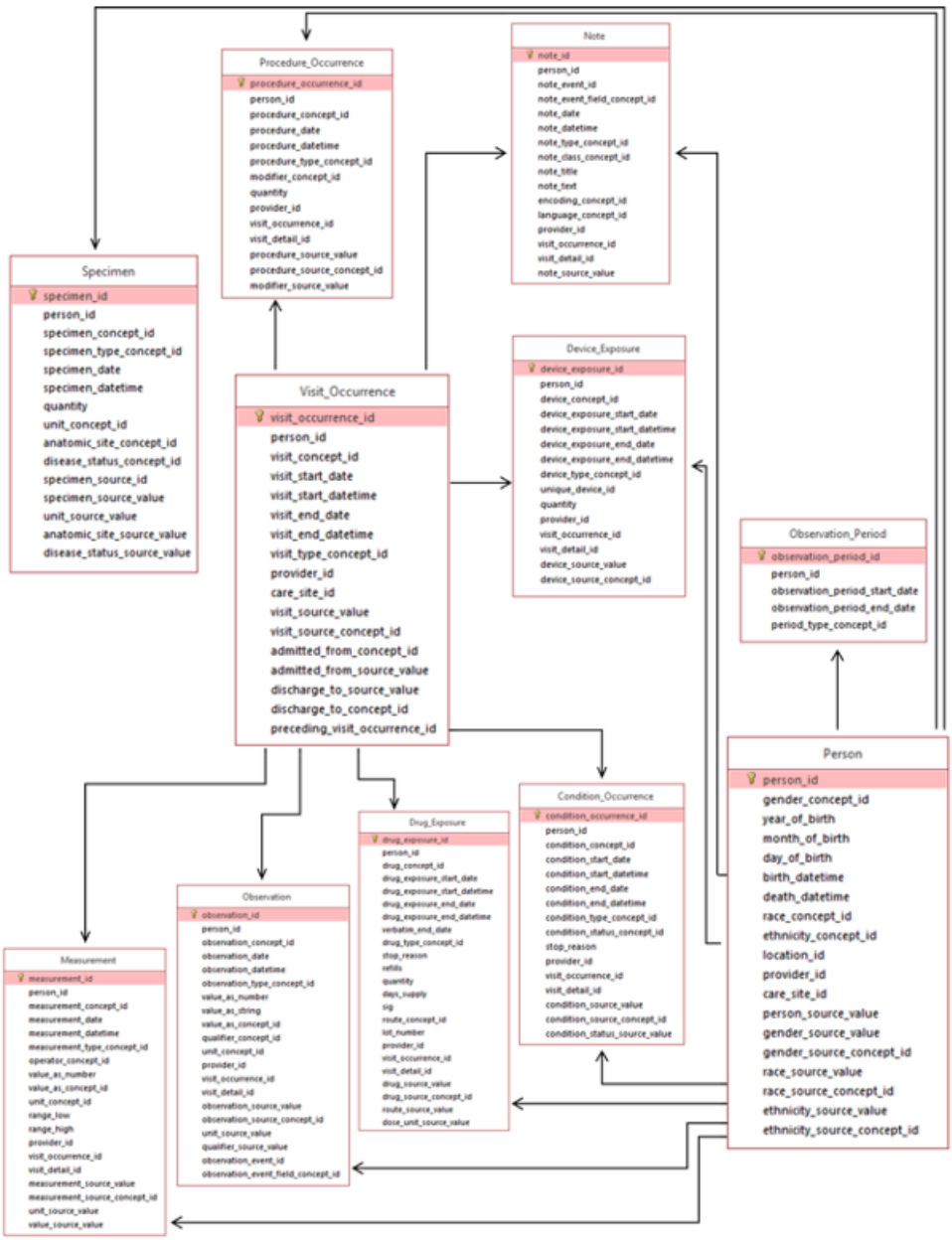

Figure 2

4. Generating Binary Feature Vectors From Concept IDs

Each distinct clinical or healthcare-related concept defined by the standardized vocabulary of the OMOP Common Data Model is encoded as a unique concept id. Under this schema, the core clinical information contained in each record of the selected standardized clinical data tables can be represented by these ids.

Each record in the *Visit_Occurrence* table encodes the type of clinical encounter as a concept id in the *visit_concept_id column*, each record in the *Condition_Occurrence* table encodes the observed disease or medical condition as a concept id in the *condition_concept_id* column, each record in the *Procedure_Occurrence* table encodes the procedure ordered or carried by as a concept id in the *procedure_concept_id* column, and each record in the *Drug_Exposure* table encodes the medication and mode of delivery as a concept id in the *drug_concept_id* column. For each of these tables, core clinical information can be extracted from the concept ids contained within a single column.

As for the clinical observations captured in the *Observation* table, the core clinical information of each record can be described by a pair of concept ids: the *observation_concept_id* which describes the type of clinical observation recorded, and the *value_as_concept_id* which encodes the value of the observation.

As we aimed to identify the subset of clinical concepts captured by the EHR which are most predictive of increased risk for POU among patients undergoing COT, we determined for each patient in the study cohort the set of concept ids he/she had been associated with by iterating through these tables. For each patient in the case group, only the records with start dates prior to his/her diagnosis date for POU were considered. The diagnosis date for each patient was identified from the *Condition_Occurrence* table by finding the earliest date on

which a concept id matching the ICD codes for POU had been recorded for that patient.

Once the set of associated concept ids were determined for each patient in the cohort, we proceeded to generate a binary matrix in which each row is a feature vector representing the concept ids associated with a particular patient.

5. Incorporating Laboratory Test Data Into Binary Feature Vectors

Conducted laboratory tests and their results are recorded in the *Measurement* standardized clinical data table under the OMOP Common Data Model. As the domain of possible results vary across tests, their results may either be encoded as a concept id or as a numerical value. When the result is directly recorded as a numerical value, the lower and upper bound of normal values is stored alongside it, enabling us to determine whether or not the result can be considered normal.

Because we seeked to assess the validity of our data processing and modelling techniques in identifying the factors most predictive of increased risk for POU, we decided to extract the results of three clinical tests from the *Measurement* table: the presence of Buprenorphine in urine, the International Normalized Ratio (INR), and the presence of Diazepam in urine.

Buprenorphine is an opioid commonly prescribed in the treatment of opioid addiction and therefore has a direct clinical association with suspected POU[14]. We identified 80 patients in our study cohort who had their urine tested for the presence of Buprenorphine. Since all 80 patients had tested negative, we proceeded to only encode whether or not this test was carried out as a feature for each patient.

The International Normalized Ratio (INR) is a measure of the time required

for the blood to clot . Although the ordering of this test is correlated with chronic disease [15], it has no direct clinical association with POU. We identified 2089 patients who had this test ordered then proceeded to encode it as an additional feature.

Each record of INR consists of a numerical test result as well as the lower and upper bound of expected normal values. A record is categorized as an abnormal result if its numerical test result falls outside of this range. Because each patient may had had multiple INR tests carried out over the course of his/her observation period, we further categorized each patient as having abnormal INR if more than 5% of these tests yielded abnormal results. Under this definition, 1464 patients were identified as having abnormal results. We also encoded this categorization as an addition feature.

Diazepam is a drug used to treat anxiety disorders or alcohol withdrawal and may be correlated with POU [16]. We identified 73 patients from the *Measurement* table who had been tested for the presence of Diazepam in their urine and encoded this as an additional feature.

To account for the possibility of having multiple tests conducted over the course of a patient's observation period, we also categorized each patient as having Diazepam detected in his/her urine if more than 5% of these tests yielded positive results. Only 1 patient in the cohort was identified to fulfill this criteria.

This set of features identified from the *Measurement* table were then concatenated with the binary feature vector previously generated for each patient. This resulted in a binary feature matrix with dimensions of 2733 x 31420.

6. Dimensionality Reduction

Given the high dimension of unique features, our dataset would be computationally slow and complex to interpret. As a result, we decided to select a proper subset to reduce complexity and overfitting [17]. One of the most common feature selection methods for a set of categorical features to be used for binary classification is to conduct a series of statistical tests to determine the likelihood of each feature variable being independent of the classification. If the feature variable is highly-likely to be independent, then it should be removed from the dataset.

Pearson's Chi-square Test is intended to test for independence between categorical variables. The scikit-learn machine library offers an implementation of the chi-squared test in the chi2() function. We conducted this test on each feature separately between the case and control groups, thereby obtaining a p-value for each feature. We then proceeded to remove features with a p-value below the threshold of 0.05.

This process resulted in the removal of 28,560 (90.9%) features and 2,860 (9.1%) features being retained. This technique identified a manageable subset of features which, when considered individually, have the highest probability of being correlated with increased risk of developing POU. However, it should be noted that this process did not consider interaction terms between features which may also have predictive power.

7. Balancing the Data

It is known that prediction models for classification will be biased in favor of the majority class if the data is imbalanced. Moreover, the bias is magnified for high-dimensional data, for which the number of features greatly exceeds the number of samples [18]. Because our dataset fulfills this criteria, we proceeded to use oversampling to adjust the class distribution such that we could simulate a class-balanced dataset. The Synthetic Minority Oversampling Technique (SMOTE) algorithm has been used in many studies, ranging from prediction of breast cancer to mRNA gene prediction [19, 20] and has been demonstrated to be a reliable method of balancing data.

Therefore, we first split our data set into a training set consisting of 2189 individuals (80%) randomly sampled from our study cohort and a corresponding testing set consisting of 547 individuals (20%). We then applied the SMOTE algorithm on the training set to ensure that none of the information in the testing set is being used to create synthetic observations, thereby maintaining the generalizability of our results. This process generated a training set containing 1845 samples of each class.

8. Baseline Logistic Model

Logistic regression is one of the regression models used most frequently in medical research and can be utilized to describe and test the relationships between a binary outcome and multiple potentially-predictive variables [21] . As a result, we aimed to train a logistic regressor using the subset of features selected through the chi2 tests in order to gain a deeper understanding of the predictive power of each feature.

We performed logistic regression on the training set with 2860 features to evaluate its ability to predict the development of POU from this set of features extracted from clinical encounters over the course of COT. This regressor achieved an accuracy of 0.852 on the testing set. The Receiver Operating Characteristic (ROC) curve of this model is shown in Figure 3 and the corresponding Area Under Curve (AUC) was calculated. AUC is a measure of the model's discriminatory ability, with a greater value representing a stronger ability in distinguishing between the two possible outcomes. An AUC value of 0.50 indicates that the model performs no better than predicting randomly, while a value of 1.00 corresponds to perfect prediction. The logistic regressor we trained reported an AUC of 0.796.

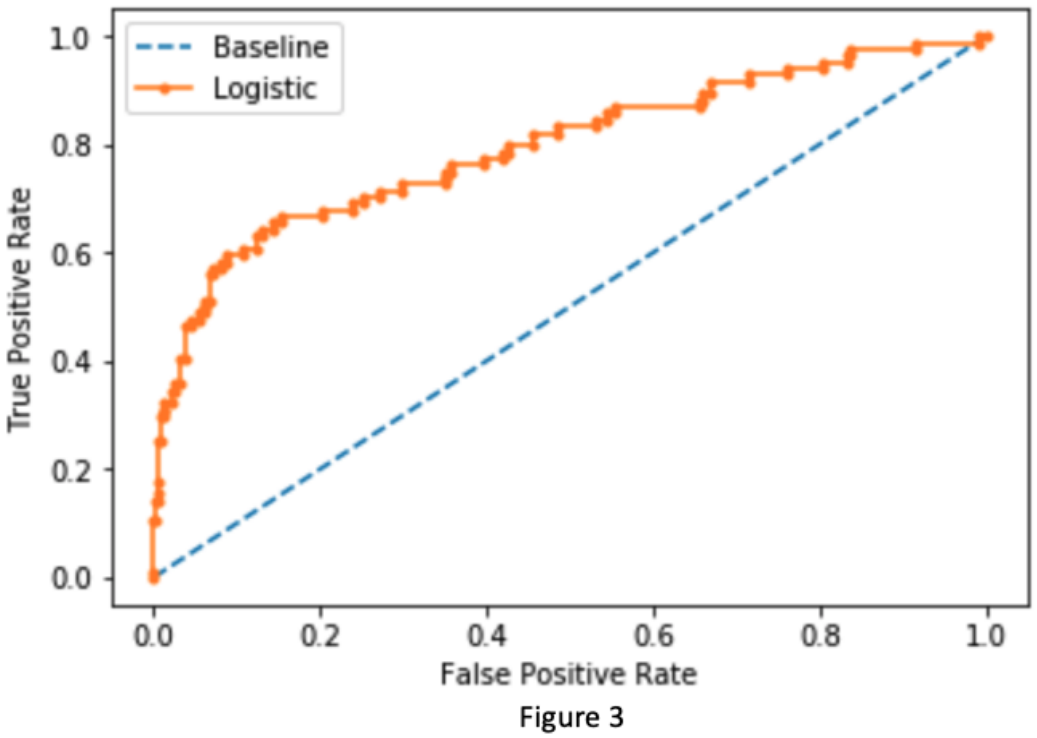

Figure 3

As this logistic regressor was fit to a feature set consisting entirely of binary categorical variables in an effort to predict a binary outcome, the resulting coefficient of each feature can be interpreted as the change in likelihood, on a log scale, of a patient developing POU as a result of having the feature. A positive coefficient indicates an increase in likelihood while a negative coefficient corresponds to a decrease in likelihood.

9. Using Recursive Feature Elimination for Feature Selection

Recursive Feature Elimination (RFE) is a feature selection method that iteratively fits a model and removes irrelevant features until the specified number of features has been reached. Over each iteration, the importance of features are ranked by the model's coefficients. Then, the features deemed least important are recursively eliminated [22]. The optimal number of features is then determined by using cross-validation to approximate the score achieved by different feature subsets.

We applied RFECV that is available in the scikit-learn machine learning library to approximate the ideal subset of features that should be included in our model. Due to time and computational constraints, we configured the algorithm to remove features in sets of size 50 and set the minimum number of features to 50 as well. Figure 4 shows the resulting RFECV curve. It is apparent that the performance achieved by logistic regression begins to plateau after approximately 1000 features. The RFECV algorithm determined that the optimal number of features to include training a logistic regressor is 1150.

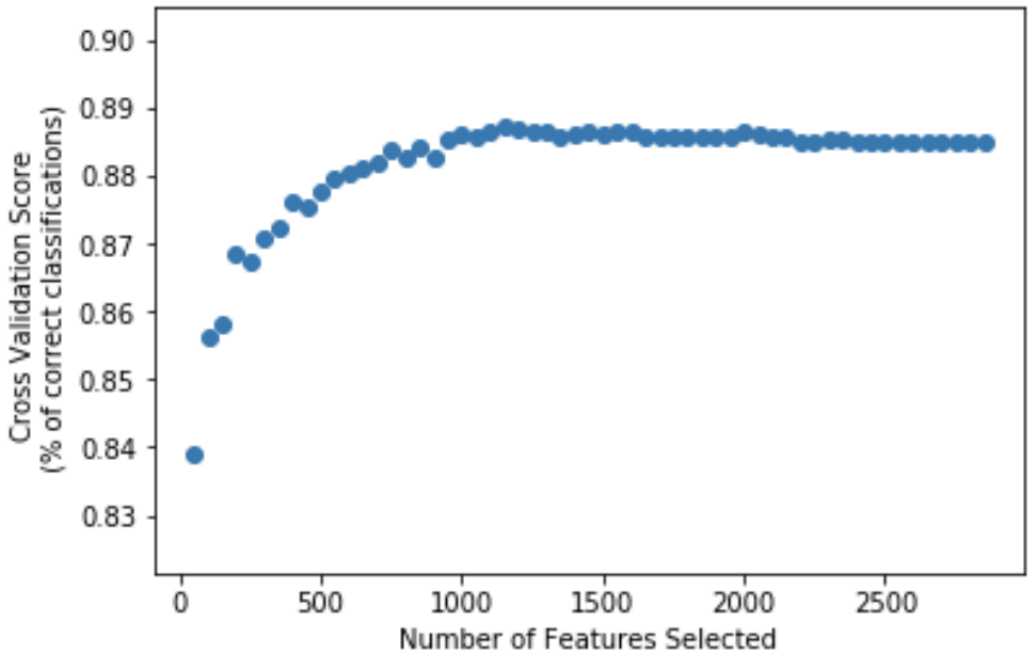

Figure 4

10. Logistic Regression Using RFECV-Selected Features

We proceeded to train a logistic regressor with the 1150 features selected by RFECV. The resulting Receiver Operating Characteristic (ROC) curve is plotted in Figure 5, with a corresponding AUC of 0.793, compared to an AUC of 0.796 achieved by the baseline model.

The coefficient associated with each feature resulting from this logistic regression is indicative of the feature's power in predicting the development of POU from our study cohort of COT patients.

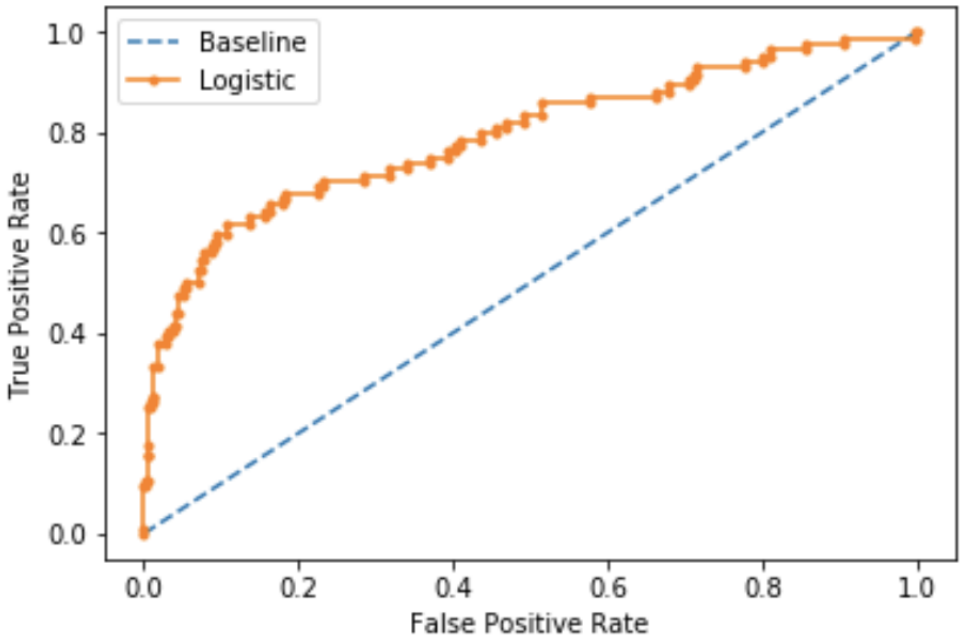
Baseline
Logistic
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0

Figure 5

CHAPTER 4

RESULTS

The logistic regressor trained using RFECV-selected features associated 656 features with positive coefficients, thereby identifying them as predictive of increased risk for POU. Of the 656 features, 313 features were concepts extracted from the *Condition_Occurrence* standardized clinical data table, 184 features were concepts extracted from the *Drug_Exposure* table, 135 features were concepts extracted from the *Procedure_Occurrence table*, and 21 features were concepts extracted from the *Observation* table.

History of emergency room and inpatient visits, extracted from the Visit table, was also identified as a predictor. Testing for the presence of Buprenorphine, a drug with a direct clinical association with POU, and testing for Diazepam, a medication used to treat anxiety disorders and alcohol withdrawal, were also identified by the logistic regressor as being positively-correlated with increased risk for POU. As expected, the features representing INR testing had not been retained.

The distribution of these features, identified as predictors of increased risk for POU, among the concept domains specified by the OMOP standardized vocabularies are shown in Figure 6 below.

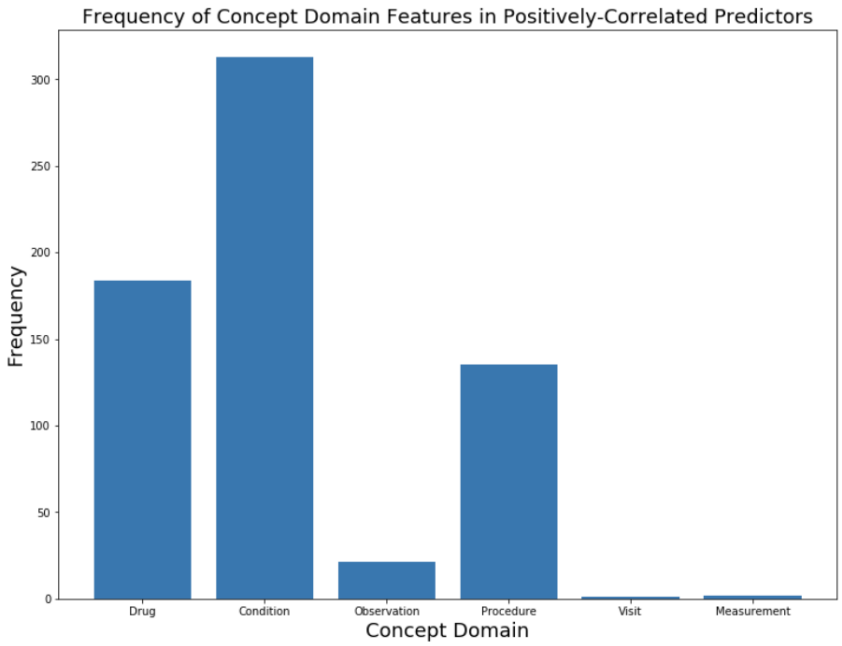

Figure 6

The 20 features with the most predictive power as determined by logistic regression include, in descending order:

1. Amylases 20000 UNT / Endopeptidases 25000 UNT / Lipase 4000 UNT Oral Capsule,
2. Lumbosacral spondylosis with radiculopathy,
3. abacavir 20 MG/ML Oral Solution [Ziagen],
4. Roflumilast 0.5 MG Oral Tablet,
5. Acetaminophen 325 MG / Oxycodone Hydrochloride 5 MG Oral Tablet [Roxicet],
6. Patient admits to alcohol use,
7. Secondary spontaneous pneumothorax,
8. Infection and inflammation associated with indwelling urinary catheter,

9. Lumbosacral spondylosis without myelopathy,
10. Alprazolam 2 MG Oral Tablet,
11. Removal of other device from thorax,
12. Diagnostic mammography, including computer-aided detection (CAD) when performed; bilateral,
13. Non megaloblastic anemia associated with nutritional deficiency,
14. Spondylosis,
15. Microsurgical techniques, requiring use of operating microscope,
16. Hospital discharge day management; 30 minutes or less,
17. Carcinoma in situ of female genital organ,
18. Lung field abnormal,
19. Nephrocalcinosis, and
20. Computed tomography, abdomen and pelvis; with contrast material(s).

The value of their associated coefficients are plotted in the bar graph below. Each bar has been colored according to its feature's OMOP concept domain.

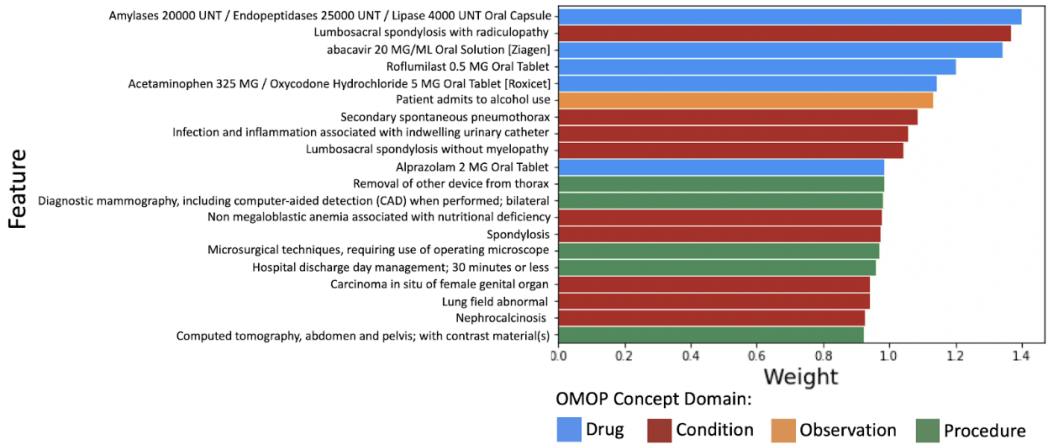

Figure 7

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

The AUC of 0.793 achieved by the logistic regressor trained on the 1150 features selected by RFECV, when compared to the AUC of 0.796 achieved by the baseline logistic regressor trained on all 2860 features extracted from the EHR, demonstrates that the data processing and modelling techniques we proposed is capable of capturing then identifying the subset of features most predictive of increased POU risk in our study cohort of patients undergoing COT.

By examining the distribution of these predictors among the concept domains specified by the OMOP standardized vocabularies, it is apparent that the medical conditions suggested by a patient's symptoms, the drugs prescribed to a patient, and the medical procedures ordered for a patient captured by the EHR can be leveraged to detect increased risk of developing POU. Furthermore, the presence of the patient admitting to alcohol use among the top 20 features with the most predictive power indicates that observations made during clinical encounters hold significant predictive value as well.

5.2 Future Work

Our results indicate that individual patient characteristics extractable from EHR data generated through clinical encounters can be leveraged by simple classification models to identify increased risk of developing problem opioid use among patients undergoing chronic opioid therapy.

Furthermore, by utilizing techniques such as the Pearson's Chi-square Test, RFECV, and examining the feature coefficients of logistic regressors, the optimal subset of features required to predict future development of POU among a cohort of patients undergoing COT can be determined. However, since this study was conducted on a sample patients of limited size at Weill Cornell Medicine, the generalizability of the identified predictors to other patient populations remains to be explored.

On the other hand, the limitations of the techniques we employed to identify features with the most predictive power are directions for future research.

Firstly, while conducting the Pearson's Chi-square Test sequentially on individual features successfully identifies a manageable subset of variables which, when considered individually, have the highest probability of being correlated with increased risk of developing POU, it fails to consider collinear interaction terms between variables which may have greater predictive power. This is especially significant when extracting predictors from EHR data because a single medical concept associated with a patient may be captured in multiple ways depending on the nature of clinical encounters. Therefore, alternative statistical tests should be considered to increase the power of identified predictors.

Secondly, due to limitations in time and available computational power, RFECV was conducted on our dataset in a stepwise manner in which groups

of 50 features were removed at a time. As a result, the algorithm will generate the optimal feature set by finding a local maximum of model performance, instead of an ideal global maximum. Therefore, the optimal number of features generated by RFECV may not be the best performing subset.

Lastly, it has been shown that there is a serious and increasing risk that naive use of analytical techniques without a full understanding of the complexities and limitations of EHR data will result in biased or incorrect medical findings [23]. Considering that our EHR data is observational databases, the data itself might contain noise. For example, the date associated with a ICD code for POU is when the physician made the diagnosis, not when the patient first developed POU. Similarly, a patient might have an abnormal INR value at some point; however, it will never be known unless a physician orders the laboratory test. Consequently, features with the most predictive power generated by the model might consist of biases. Therefore, we might consider conducting a manual chart review of each patient in our cohort to better grasp the causation between risk factors and target outcome and validate the results generated.

CHAPTER 6

ACKNOWLEDGEMENT

Our deep gratitude goes to Dr. Curtis Leland Cole for his continuous support over the course of this Specialization Project. His deep expertise, enthusiasm, and patience proved instrumental as he guided us through the research process. In addition, our appreciation goes to Evan Sholle for offering us key advice and connecting us to needed resources. His knowledge and prompt suggestions have enabled us to solve challenges effectively and complete this project.

BIBLIOGRAPHY

1. Deborah Dowell, Tamara M. Haegerich, and Roger Chou. 2016. CDC Guideline for Prescribing Opioids for Chronic Pain—United States, 2016. Jama315, 15 (March 2016), 1–49. http://dx.doi.org/10.1001/jama.2016.1464

2. Elizabeth M. Olivia et al. 2017. Development and Applications of the Veterans Health Administration's Stratification Tool for Opioid Risk Mitigation (STORM) to Improve Opioid Safety and Prevent Overdose and Suicide. Psychological Services 14, 1 (2017), 34–49. http://dx.doi.org/10.1037/ser0000099.supp

3. Kathleen T. Brady, Jenna L. Mccauley, and Sudie E. Back. 2015. Prescription Opioid Misuse, Abuse, and Treatment in the United States: An Update. American Journal of Psychiatry173, 1 (September 2015), 18–26. http://dx.doi.org/10.1176/appi.ajp.2015.15020262

4. Timothy R. Hylan et al. 2015. Automated Prediction of Risk for Problem Opioid Use in a Primary Care Setting. The Journal of Pain 16, 4 (2015), 380–387. http://dx.doi.org/10.1016/j.jpain.2015.01.011

5. Roy E. Palmer et al.2015. The prevalence of problem opioid use in patients receiving chronic opioid therapy. Pain156, 7 (2015), 1208–1214. http://dx.doi.org/10.1097/j.pain.0000000000000145

6. David S. Carrell . 2015. Using natural language processing to identify problem usage of prescription opioids. International Journal of Medical Informatics(September 2015), 1057-1064. http://dx.doi.org/10.1016/j.ijmedinf.2015.09.002

7. Mark R. Jones, Omar Viswanath, Jacquelin Peck, Alan D. Kaye, Jatinder S. Gill, and Thomas T. Simopoulos. 2018. A Brief History of the Opioid Epidemic and Strategies for Pain Medicine. Pain and Therapy7, 1 (2018), 13–21. http://dx.doi.org/10.1007/s40122-018-0097-6

8. Chou, Roger et al. Clinical Guidelines for the Use of Chronic Opioid Therapy in Chronic Noncancer Pain.The Journal of Pain, Volume 10, Issue 2, 113 - 130.e22 https://doi.org/10.1016/j.jpain.2008.10.008

9. Mark J. Edlund, Bradley C. Martin, Ming-Yu Fan, Andrea Devries, Jennifer B. Braden, and Mark D. Sullivan. 2010. Risks for opioid abuse and dependence among recipients of chronic opioid therapy: Results from the TROUP Study. Drug and Alcohol Dependence112, 1-2 (2010), 90–98. http://dx.doi.org/10.1016/j.drugalcdep.2010.05.017

10. Elizabeth T. Masters et al.2018. Natural Language Processing–Identified Problem Opioid Use and Its Associated Health Care Costs. Journal of Pain Palliative Care Pharmacotherapy32, 2-3 (2018), 106–115. http://dx.doi.org/10.1080/15360288.2018.1488794

11. David Atkins, Joel Kupersmith, and Seth Eisen. 2010. The Veterans Affairs Experience: Comparative Effectiveness Research In A Large Health System. Health Affairs29, 10 (2010), 1906–1912.
http://dx.doi.org/10.1377/hlthaff.2010.0680

12. Jenna M Reps, Martijn J Schuemie, Marc A Suchard, Patrick B Ryan, Peter R Rijnbeek, Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data, Journal of the American Medical Informatics Association, Volume 25, Issue 8, August 2018, Pages 969–975, https://doi.org/10.1093/jamia/ocy032

13. Observational Health Data Sciences and Informatics. (n.d.). Retrieved from https://www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm:single-page.

14. LOINC 3414-0 - Buprenorphine [Presence] in Urine. (n.d.). Retrieved from https://loinc.org/3414-0/.

15. LOINC 3414-0 - Buprenorphine [Presence] in Urine. (n.d.). Retrieved from

https://loinc.org/3414-0/.

16. LOINC 3414-0 - Buprenorphine [Presence] in Urine. (n.d.). Retrieved from https://loinc.org/3414-0/.

17. Kaushik, S., Saurav. (2019, June 24). Feature Selection methods with example (Variable selection methods). Retrieved from https://www.analyticsvidhya.com/blog/2016/12/introduction- to-feature -selection-methods-with-an-example-or-how-to-select-the-right-variables/

18. Blagus, R., Lusa, L. SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics 14, 106 (2013) doi:10.1186/1471-2105-14-106 19. Fallahi, Amir Jafari, Shahram. (2011). An Expert System for Detection of Breast Cancer Using Data Preprocessing and Bayesian Network. Int J Adv Sci Technol. 34.

20. Xiao, J., Tang, X., Li, Y. et al. Identification of microRNA precursors based on random forest with network-level representation method of stem-loop structure. BMC Bioinformatics 12, 165 (2011) doi:10.1186/1471-2105-12-165

21. Hess, A. S., Hess, J. R. (2019). Logistic regression. Transfusion, 59(7), 2197–2198. doi: 10.1111/trf.15406

22. Recursive Feature Elimination. (n.d.). Retrieved from https://www.scikit-yb.org/en/latest/api/model_selection/rfecv.html.

23. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. (2018). Bmj. doi: 10.1136/bmj.k4416