

Iowa Initiative for Artificial Intelligence

Final Report

Project title:	Identifying Risk Factors to Predict Adverse Outcomes Related to Substance Use	
Principal Investigator:	Anna Merrill	
Prepared by (IIAI):	Nam H Le	
Other investigators:	Juan Pablo Hourcade, Brandi Janssen, Heath Davis, Ryan Carnahan	
Date:	June 28, 2021	
Were specific aims fulfilled:	No	
Readiness for extramural proposal?	No	
If yes ... Planned submission date		
Funding agency		
Grant mechanism		
If no ... Why not? What went wrong?	Insufficient epidemiological study design as well as computational limitation.	

Brief summary of accomplished results:

The specific aims were not fulfilled during the lifetime of the project. The project faced multiple drawbacks:

First, there were multiple unexpected challenges with data format and access. It was not until November 2020 that all research team members were able to access the TriNetX data on the ICTS virtual server. An update to the TriNetX health research network in March 2021 provided us with the ability to differentiate between different types of encounters with the healthcare system (e.g., emergency, inpatient, ambulatory). In order to utilize this information as an input variable, an updated TriNetX data set was retrieved, which, unexpectedly, also had differences in Medication and Procedure data formats that had to be accommodated.

Second, due to the abovementioned issues and also due to insufficient clarity of study design, the process of data cleaning and preprocessing was not accomplished until April 2021. One significant, ongoing challenge is the definition of controls in this data set. Incidence density sampling with replacement is considered the gold standard control sampling method, but has not yet been able to be applied to this data set.

Finally, as is currently required by UIHC, the data set had to reside behind a highly secure firewall in a computer with limited number of CPUs, RAM, and no GPUs. The initial computational resources requested for this project were not sufficient for the desired complex analysis in decision tree-based classifiers and DL methods, thereby hindering progress. Throughout the project lifetime, multiple requests to move the data source to a sufficiently powered server were not granted.

Despite these drawbacks, the team managed to move their research forward and accomplished the following:

- Understanding of the TriNetX data and its tables, including how to use and integrate the Patient, Encounter, Vital Signs, Diagnosis, Procedure, Medication, and Lab Result Tables.
- Selection of encounters per case by choosing different time windows before the first overdose.
- Formulation of grouping strategy for Medication (i.e., RxNorm), Diagnosis (i.e., ICD-9 and ICD-10), Procedure (i.e., CPT), and Lab result (i.e., LOINC) codes
- Creation of strategy for summarizing encounter-level information for patient-level analysis

This progress makes next steps easier and feasible – once data can actually be processed on a sufficiently powerful GPU-enabled computer to which the sensitive data can be transferred for processing.

After the end of the project pilot, one of the project Co-Investigators will work with a graduate student to continue progress toward the specific aims outlined.

Research report:

Aims: To understand which individuals are at greatest risk of adverse outcomes related to substance use, such as overdose, trauma, and socioeconomic distress.

The primary specific aim of the proposed research was to use demographic information (e.g., sex, age, race, ethnicity), encounters with the healthcare system, prescription medications (e.g., opioid analgesics, benzodiazepines), concurrent alcohol use disorder, and associated diagnoses to predict adverse outcomes, such as overdose, trauma, injury, and abuse, in individuals with documented substance use disorder.

The secondary specific aim was to use demographic information (e.g., sex, age, race, ethnicity), encounters with the healthcare system, concurrent alcohol use disorder, and associated diagnoses to predict adverse outcomes, such as new diagnoses of substance use disorder, overdose, trauma, injury, and abuse, in individuals prescribed opioid analgesics and/or benzodiazepines.

Data:

Data to support the proposed research were available through the TriNetX health research network supported by the Institute for Clinical and Translational Science.

AI/ML and Data Analysis Approach:

a. Identifying outcome variable:

A number of ICD-10 Diagnosis codes has been identified to be used as selection criteria for the first outcome variable of interest in both specific aims – overdose:

- Opium (T40.0X1 A, T40.0X2, T40.0X3, T40.0X4 A,D)
- Heroin (T40.1 X1, X2, X3, X4)
- Other opioids (T40.2X1, T40.2X2, T40.2X3, T40.2X4)

- Methadone (T40.3X1, T40.3X2, T40.3X3, T40.3X4)
- Other synthetic narcotics (T40.4X1, T40.4X2, T40.4X3, T40.4X4)
- Cocaine (T40.5X1, T40.5X2, T40.5X3, T40.5X4)
- Unspecified narcotics (T40.601, T40.602, T40.603, T40.604)
- Other narcotics (T40.691, T40.692, T40.693, T40.694)
- Cannabis (T40.7X1, T40.7X2, T40.7X3, T40.7X4)
- LSD (T40.8)
- Unspecified psychodysleptics/hallucinogens (T40.901, T40.902, T40.903, T40.904)
- Other psychodysleptics/hallucinogens (T40.991, T40.992, T40.993, T40.994)
- Barbiturates (T42.3X1, T42.3X2, T42.3X3, T42.3X4)
- Benzodiazepines (T42.4X1, T42.4X2, T42.4X3, T42.4X4)
- Amphetamines (T43.621, T43.622, T43.623, T43.624)
- Methylphenidate (T43.631, T43.632, T43.633, T43.634)
- Ecstasy (T43.64)

b. Defining cases/controls:

Among those who experienced overdose, to become a case, they need to have at least one encounter happening before the first recorded overdose. A time window is used to select prior encounters happening before the first overdose to ensure that sufficient information about an individual could be gathered prior to that individual experiencing overdose. Table 1 shows the effects of choosing a time window of 1, 2, 3, 5, and 7 year on the resulting number of selected patients and encounters for cases. Based on this analysis, only patients with at least one encounter within 2 years of experiencing overdose will be included as a case.

Table 1. Number of patients and encounters using different time windows

	encounter_count	patient_count
aim1_diagnosis	9,897,675	24,081
aim1_diagnosis_icd10	5,088,473	x
aim1_diagnosis_icd10_overdose		
aim1_diagnosis_icd10_overdose_A	937	778
aim1_encounter_icd9&10_before_overdose_1y	10,760	535
aim1_encounter_icd9&10_before_overdose_2y	18,592	578
aim1_encounter_icd9&10_before_overdose_3y	25,742	606
aim1_encounter_icd9&10_before_overdose_5y	36,329	631
aim1_encounter_icd9&10_before_overdose_7y	45,664	648
aim2_diagnosis	76,634,077	338,488
aim2_diagnosis_icd10	40,295,643	x
aim2_diagnosis_icd10_overdose		

aim2_diagnosis_icd10_overdose_A	1,315	1121
aim1_encounter_icd9&10_before_overdose_1y	16,040	797
aim1_encounter_icd9&10_before_overdose_2y	27,745	861
aim1_encounter_icd9&10_before_overdose_3y	37,762	900
aim1_encounter_icd9&10_before_overdose_5y	53,066	937
aim1_encounter_icd9&10_before_overdose_7y	65,798	959

c. Variables of cases/control:

Since the TriNetX data provides predominantly encounter-level information and the desired analysis is on the patient-level, a strategy to summarize information from encounters was proposed to generate features for cases.

Table 2. Features for cases and how to summarize from encounters

Feature Name	Type	Features	Data file	How to summarize from encounters
Patient characteristics	Continuous	Age	file = patient.csv, column = year_of_birth (some are missing due to Protected status), will have to calculate age at time of outcome based on encounter year	Use value from encounter prior to encounter where outcome occurs
Patient characteristics	Continuous	BMI	file = vitals_signs.csv, column = code (39156-5 is BMI)	Use value from encounter prior to encounter where outcome occurs
Patient characteristics	Categorical	Sex	file = patient.csv, column = sex, possible values = F, M, Unknown	Use value from encounter prior to encounter where outcome occurs
Patient characteristics	Categorical	Race	file = patient.csv, column = race, possible values = American Indian or Alaska Native, Asian, Black or African American, Unknown, White	Use value from encounter prior to encounter where outcome occurs
Patient characteristics	Categorical	Ethnicity	file = patient.csv, column = ethnicity, possible values = Hispanic or Latino, Not Hispanic or Latino, Unknown	Use value from encounter prior to encounter where outcome occurs
Patient characteristics	Categorical	Marital Status	file = patient.csv, column = marital_status, possible values = Married, Single, Unknown	Use value from encounter prior to encounter where outcome occurs
Healthcare system encounters	Continuous	Yearly encounters	file = encounter.csv, column = type, possible values =	# encounters per type prior to outcome (using previous

		(by type and overall)	AMB (Ambulatory), EMER (Emergency), IMP (Inpatient), OBSENC (Observation), Unknown, Total	encounters leading up to encounter w/ outcome) -- 6 months before, 1 year before, 2 years before
Comorbid conditions/diagnosis	Categorical	Diagnosis groups from previous encounters (CCS Level 2 for ICD-9 and ICD-10)	file = Diagnosis_Labeled_CCS.csv, column = L2DCCS1	Type of diagnosis within certain time frame prior to outcome (using previous encounters leading up to encounter w/ outcome) -- 6 months before, 1 year before, 2 years before
Procedures	Categorical	Procedure groups from previous encounters (CCS Level 1 for CPT)	file = Procedure_Labeled.csv, column = cpt_ccs1	Type of procedure within certain time frame prior to outcome (using previous encounters leading up to encounter w/ outcome) -- 6 months before, 1 year before, 2 years before
Medications	Continuous	# days on medication type per encounter type	file = medication_ingredient.csv, column = code (link to Labs-Meds file)	# medication type days per encounter type prior to outcome (using all previous encounters leading up to encounter w/ outcome) -- 6 months before, 1 year before, 2 years before
Laboratory results	Continuous	Yearly positive lab results (by type and overall)	file = lab_result.csv, column = code (link to Labs-Meds file)	# positive lab result type/year prior to outcome (using all previous encounters leading up to encounter w/ outcome) -- 6 months before, 1 year before, 2 years before

d. AI approach:

Not applicable since the strategy to select controls and its features has not been finalized.

Ideas/aims for future extramural project:

Not applicable

Publications resulting from project:

None to date