ONC 2023

ANNUAL MEETING

Breakout Session: Leading Edge Acceleration Projects: Cutting Edge Health IT Tools for Healthcare and Healthcare Research

@ONC_HealthIT

Share your content on X and don't forget to use the hashtag

#ONC2023

MedStar Health

It's how we treat people.

FHIR Factories: Cutting Edge Health IT Tools for Scaling Health Research

Kristen Miller, DrPH, MSL, MSPH, CPPS
Sr. Scientific Director, MedStar Health National Center for Human Factors in Healthcare
Associate Professor of Emergency Medicine, Georgetown University School of Medicine
Affiliate Faculty, Georgetown Innovation Center for Biomedical Informatics

Background: ONC LEAP 2020

Area 2: Cutting Edge Health IT Tools for Scaling Health Research

- Exponential growth of EHR data but there are challenges in capitalizing on the value of these data due to difficulties with the data and the health IT infrastructure.
  - Challenges and barriers in tool identification
  - Lack of clarity of tool purpose
  - Poor usability/ usefulness
  - Lack of open use standards
- The ONC Policy and Development Agenda identified the need to develop tools and functions that leverage health IT infrastructure to better support research.

Research Objectives

1. Conduct a landscape analysis to explore, identify, and describe current open health IT-based tools for research.
2. Identify needs for new health IT-based tools for research that require development.
3. Address those needs through the project team's novel Fast Healthcare Interoperability Resources (FHIR) Factory platform.
4. Evaluate operational utility of the novel solution to report the potential impact, opportunities, and challenges.

Research Overview and Highlights

- Environmental Scan of Open-Source Health IT Tools
  - Scoping review to identify open-source, open standards health IT tools used for biomedical and health services research
  - Initial scan include 3,343 tools; 121 tools met inclusion criteria
  - Data extraction: licensing, languages, tool type, documentation quality, community support, longevity/pedigree, interoperability, data lifecycle, support
- Health IT Usability Evaluations
  - Evaluation included tool availability review, user-facing review, and SME review for resource requirements and system level dependencies
  - Tools were ultimately organized into 3 levels of complexity
  - Tools had generally high levels of usability but more attention must be placed on making software accessible to and useful for the broader scientific community

Research Overview and Highlights

- Stakeholder Interviews and Steering Committee Feedback
  - 16 interviews with researchers, health IT developers, health IT policy experts
  - Discussion with Steering Committee (leaders in the health IT field)
  - Identified discrepancies in the definition of open-source, challenges in adoption and promotion, the importance of community support, lack of sufficient incentives, and the need for tools to support data maintenance (i.e., data cleaning, enrichment, and transformation)
- Horizon Scan
  - We identified 16 clusters representing technological and system innovations
  - Topic levels categories: (1) interoperability and data security; (2) data types; (3) public health data; and (4) unclassified.

Research Overview and Highlights

Comprehensive List of Critical Needs

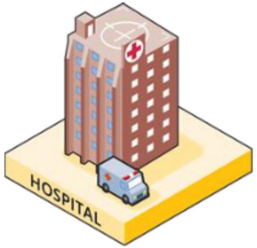
1. Need to support innovation in health IT research
2. Workforce challenges
3. Health IT standards/interoperability
4. Privacy and security
5. Embracing and optimizing open-source health IT community
6. Policy
7. Technical
8. Quality/incentives
9. Governance
10. Limited FHIR capabilities
11. Public health data

Technical Overview and Highlights

- FHIR Factories Infrastructure
  - Conceptually, a data factory that leverages data in the format of HL7's FHIR standard to conduct at-scale extraction and analysis
  - Data Processing Automation (DPA) and Robotic Process Automation (RPA) to enable complex transformations of data
- Two Demonstration Cases
  - "Bugs & Drugs": Near-real time continuous antibiograms that defines drug of choice for infection
  - "Trend Engine": A generalizable trend monitoring infrastructure

FHIR FACTORY OVERVIEW

HOSPITAL

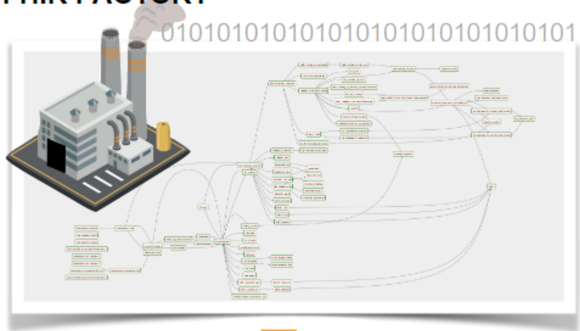


1 - EXTRACTION

Data is extracted from an EMR via BULK FHIR or INDIVIDUAL PERSON QUERY when BULK FHIR is not supported.

FHIR

FHIR FACTORY



2 - TRANSFORMATION

A published, peer reviewed medical article serves as an operational template for a FHIR FACTORY. The factory is built to reproduce all the data science in the article in an automated fashion using open-source computational orchestration infrastructure (e.g. Apache Airflow)

FHIR APP

0101
1010
0101
1010
0101
1010
0101
1010

Open Microbiol J. 2017; 11: 292–300. PMCID: PMC5688387
Published online 2017 Oct 31. PMID: 29204224
doi: 10.2174/1874285801711010292

Bacteriology and Antibiogram of Urinary Tract Infection Among Female Patients in a Tertiary Health Facility in South Eastern Nigeria

Angus N. Oli,[1,*] Vivian B. Akabueze,[1] Chijioke E. Ezeudu,[2] George U. Eleje,[3] Obiora S. Ejiofor,[4] Ifeanyichukwu U. Ezebialu,[5] Charlotte B. Oguejiofor,[3] Ifeoma M. Ekejindu,[6] George O. Emechebe,[4] and Kenneth N. Okeke[4]

3 - OUTPUT

The output of the FHIR FACTORY is both *human* readable knowledge and *computational* readable. Humans can read a medical article, whereas digital systems can leverage an API to the results of the article or access the AI model.

4 - INTEGRATION

The output of the FHIR FACTORY can be integrated into workflows through the use of SMART on FHIR apps or use of the API in 3rd party applications. The result can create the closed loop elements of a continuous Learning Healthcare System.

TREND ENGINES

FHIR

main.py — venv

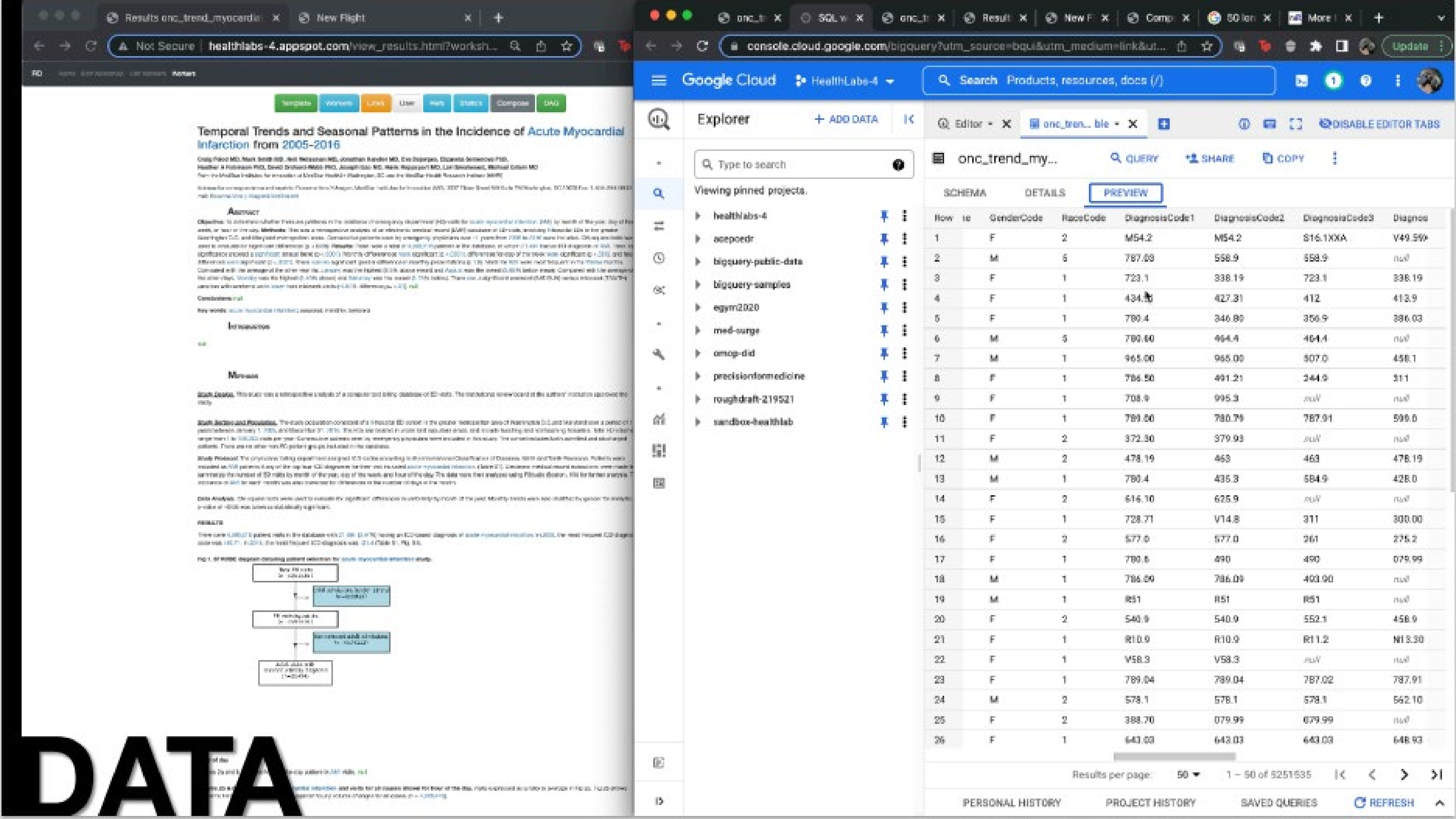
main.py M × | fhirdata.py | sandbox.py 1 | condition.py 2, M | kv.py | connections.py 2 | crypto_create_key.py | encounter.py | observ...

main.py > iterate_cmrn_global

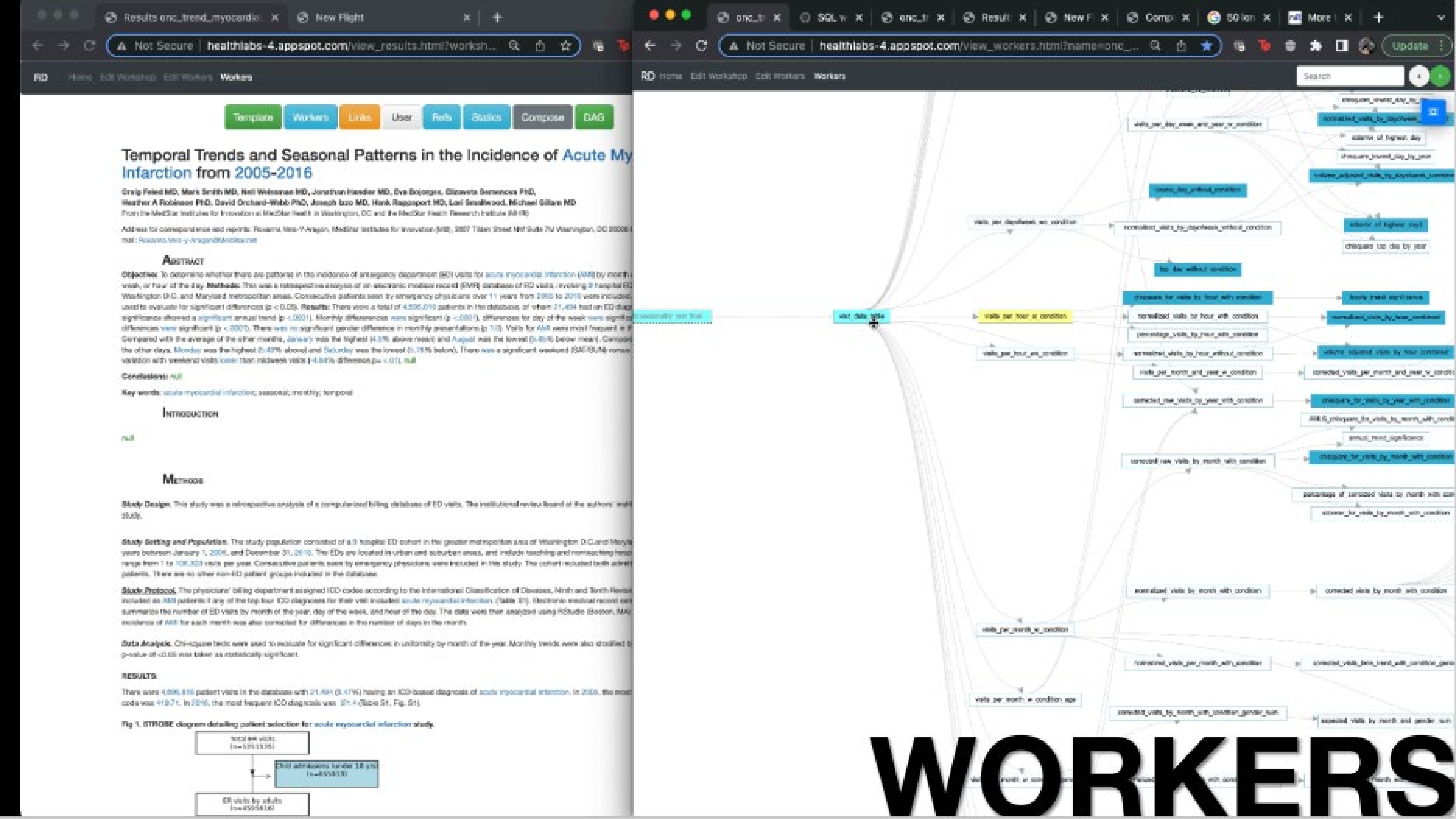
VENV
- __pycache__
- .gitignore M
- condition.py 2, M
- config.py
- connections_new.py
- connections_old.py
- connections.py 2
- crypt_key.bin U
- crypto_create_key.py
- crypto.py
- enc_b41_200rows.bin U
- enc_b41_200rows.in U
- enc_b41.bin U
- enc_b41.in U
- enc.in U
- encounter.py
- encrypted.bin U
- fhirdata.py
- hapi.py
- kv.py
- main.py M
- methods.py
- observation.py
- patient.py
- patients.in U
- sandbox.py 1
- secret_b41.in
- synthea.py
- test.bin U

OUTLINE

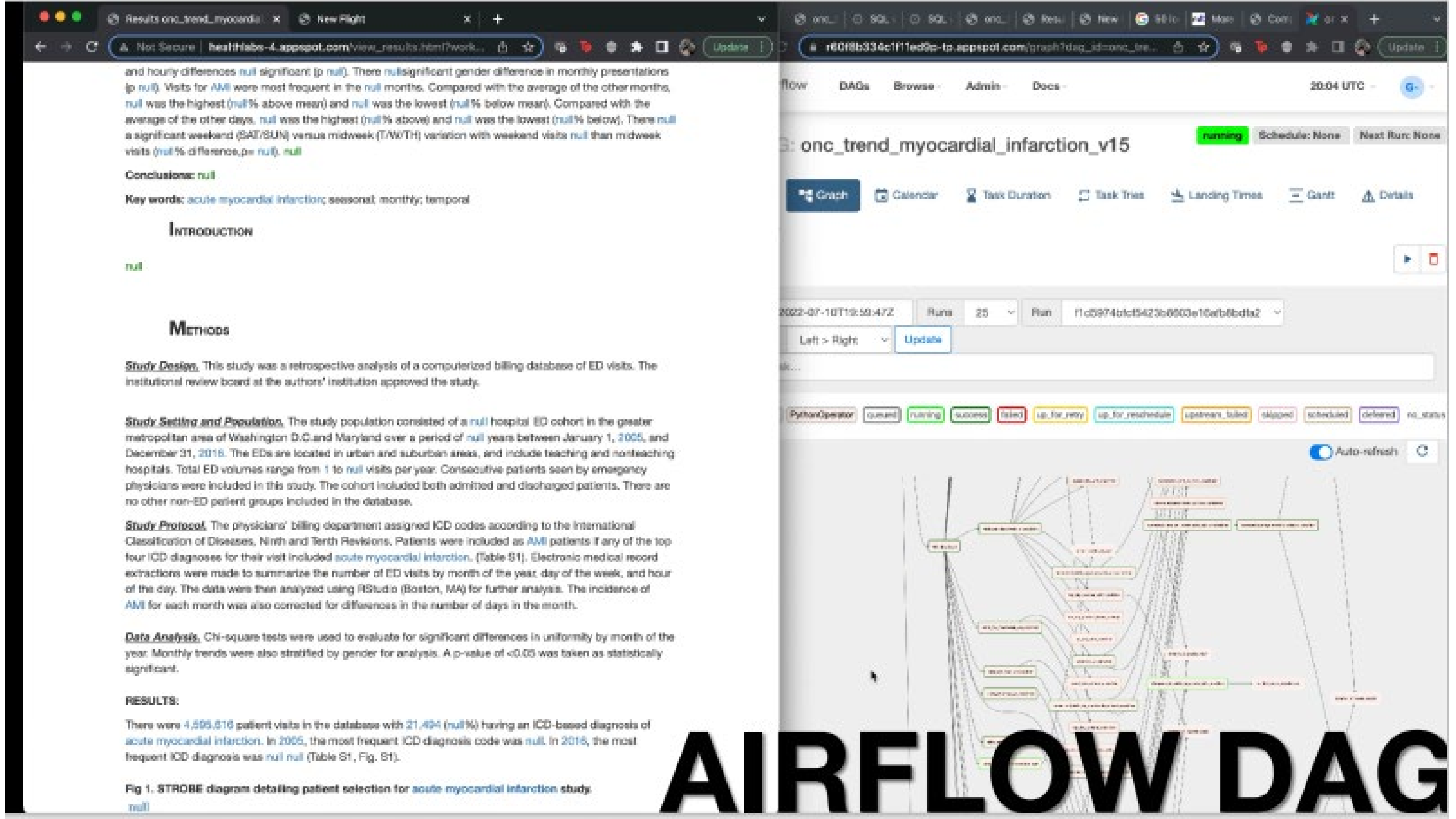
```python
        print(f"{i} complete...")

###OLD METHODs####

def iterate_cmrn(filename):
    patients = []
    conditions = []
    encounters = []
    time_start = time.perf_counter()
    bearer_start = time_start
    bearer = get_bearer_token() #REMEMBER IT ONLY LASTS 10 MINUTES
    id_list = [line.rstrip() for line in open(filename,'r')] #want to close, eventually
    for i in id_list:
        time_elapsed = (time.perf_counter() - time_start)
        bearer_time_elapsed = (time.perf_counter()-bearer_start)
        print ("%s\t%5.1f secs %5.1f secs" % (i,time_elapsed,bearer_time_elapsed))
        if bearer_time_elapsed>550:
            bearer_start=time.perf_counter()
            bearer = get_bearer_token()
        pid = get_pid(bearer,i)
        #print(pid)
        pt = make_request(bearer,'Patient','_id='+pid).json()
        try:
            patients.append(Patient(pt['entry'][0]['resource']))
            dx = make_request(bearer,'Condition','patient='+pid+'&category=encounter-diagnosis&clinical-status=active').json()
            for j in range(0,len(dx['entry'])):
                cond = Condition(dx['entry'][j])
                if cond.icd_code is not None:
                    conditions.append(cond)
                    enc = make_request(bearer,'Encounter','_id='+cond.encounter_id).json()
                    encounters.append(Encounter(enc['entry'][0]))
        except Exception as e:
            print (e)
            continue

    return patients,conditions,encounters

def encdx(pid):
    time_start = time.perf_counter()
    conditions = []
    encounters = []
    try:
        dx = make_request_global('Condition','patient='+pid+'&category=encounter-diagnosis&clinical-status=active').json()

        for j in range(0,len(dx['entry'])):
            cond = Condition(dx['entry'][j])
            if cond.icd_code is not None:
```


Results onc_trend_myocardial
New Flight
Not Secure | healthlabs-4.appspot.com/view_results.html?worksh...
Template
Workers
Lines
User
Hers
Statics
Compose
DAG
Temporal Trends and Seasonal Patterns in the Incidence of Acute Myocardial Infarction from 2005-2016
Abstract
Introduction
Methods
console.cloud.google.com/bigquery?utm_source=bqui&utm_medium=link&ut...
Google Cloud
HealthLabs-4
Search Products, resources, docs (/)
Explorer
+ ADD DATA
Type to search
Viewing pinned projects.
healthlabs-4
acepoedr
bigquery-public-data
bigquery-samples
egym2020
med-surge
omop-did
precisionformedicine
roughdraft-219521
sandbox-healthlab
Editor
onc_tren...ble
DISABLE EDITOR TABS
onc_trend_my...
QUERY
SHARE
COPY
SCHEMA
DETAILS
PREVIEW
Row | GenderCode | RaceCode | DiagnosisCode1 | DiagnosisCode2 | DiagnosisCode3 | Diagnos
1 | F | 2 | M54.2 | M54.2 | S16.1XXA | V49.59
2 | M | 5 | 787.03 | 558.9 | 558.9 | null
3 | F | 1 | 723.1 | 338.19 | 723.1 | 338.19
4 | F | 1 | 434.91 | 427.31 | 412 | 413.9
5 | F | 1 | 780.4 | 346.80 | 356.9 | 386.03
6 | M | 5 | 780.60 | 464.4 | 464.4 | null
7 | M | 1 | 965.00 | 965.00 | 307.0 | 458.1
8 | F | 1 | 786.50 | 491.21 | 244.9 | 311
9 | F | 1 | 708.9 | 995.3 | null | null
10 | F | 1 | 789.00 | 780.79 | 787.01 | 599.0
11 | F | 2 | 372.30 | 379.93 | null | null
12 | M | 2 | 478.19 | 463 | 463 | 478.19
13 | M | 1 | 780.4 | 435.3 | 584.9 | 428.0
14 | F | 2 | 616.10 | 625.9 | null | null
15 | F | 1 | 728.71 | V14.8 | 311 | 300.00
16 | F | 2 | 577.0 | 577.0 | 261 | 275.2
17 | F | 1 | 780.6 | 490 | 490 | 079.99
18 | M | 1 | 786.09 | 786.09 | 493.90 | null
19 | M | 1 | R51 | R51 | R51 | null
20 | F | 2 | 540.9 | 540.9 | 552.1 | 456.9
21 | F | 1 | R10.9 | R10.9 | R11.2 | N13.30
22 | F | 1 | V58.3 | V58.3 | null | null
23 | F | 1 | 789.04 | 789.04 | 787.02 | 787.91
24 | M | 2 | 578.1 | 578.1 | 578.1 | 562.10
25 | F | 2 | 388.70 | 079.99 | 079.99 | null
26 | F | 1 | 643.03 | 643.03 | 643.03 | 648.93
Results per page: 50
1 – 50 of 5251535
PERSONAL HISTORY
PROJECT HISTORY
SAVED QUERIES
REFRESH

DATA



Results onc_trend_myocardia
New Flight
Not Secure | healthlabs-4.appspot.com/view_results.html?worksh...
RD Home Edit Workshop Edit Workers Workers
Template Workers Links User Refs Statics Compose DAG
Temporal Trends and Seasonal Patterns in the Incidence of Acute My
Infarction from 2005-2016
Abstract
Objectives:
Methods:
Results:
Conclusions: null
Key words: acute myocardial infarction; seasonal; monthly; temporal
Introduction
null
Methods
Study Design.
Study Setting and Population.
Study Protocol.
Data Analysis.
RESULTS:
Fig 1. STROBE diagram detailing patient selection for acute myocardial infarction study.
Not Secure | healthlabs-4.appspot.com/view_workers.html?name=onc_...
RD Home Edit Workshop Edit Workers Workers
Search
Update
visits per hour w condition
WORKERS



and hourly differences null significant (p null). There nullsignificant gender difference in monthly presentations (p null). Visits for AMI were most frequent in the null months. Compared with the average of the other months, null was the highest (null% above mean) and null was the lowest (null% below mean). Compared with the average of the other days, null was the highest (null% above) and null was the lowest (null% below). There null a significant weekend (SAT/SUN) versus midweek (T/W/TH) variation with weekend visits null than midweek visits (null% difference,p= null). null

Conclusions: null

Key words: acute myocardial infarction; seasonal; monthly; temporal

Introduction

null

Methods

Study Design. This study was a retrospective analysis of a computerized billing database of ED visits. The institutional review board at the authors' institution approved the study.

Study Setting and Population. The study population consisted of a null hospital ED cohort in the greater metropolitan area of Washington D.C.and Maryland over a period of null years between January 1, 2005, and December 31, 2016. The EDs are located in urban and suburban areas, and include teaching and nonteaching hospitals. Total ED volumes range from 1 to null visits per year. Consecutive patients seen by emergency physicians were included in this study. The cohort included both admitted and discharged patients. There are no other non-ED patient groups included in the database.

Study Protocol. The physicians' billing department assigned ICD codes according to the International Classification of Diseases, Ninth and Tenth Revisions. Patients were included as AMI patients if any of the top four ICD diagnoses for their visit included acute myocardial infarction. (Table S1). Electronic medical record extractions were made to summarize the number of ED visits by month of the year, day of the week, and hour of the day. The data were then analyzed using RStudio (Boston, MA) for further analysis. The incidence of AMI for each month was also corrected for differences in the number of days in the month.

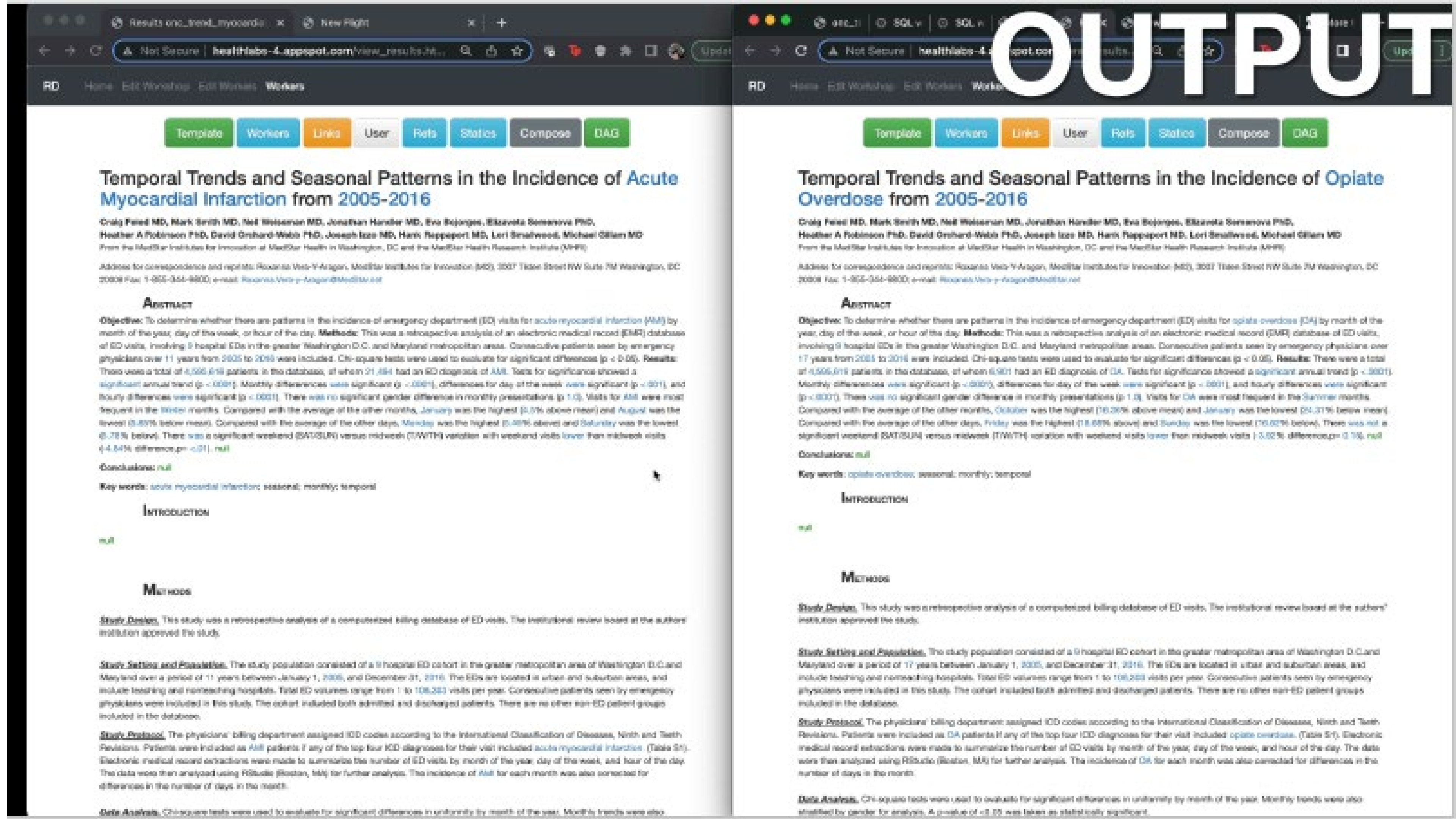
Data Analysis. Chi-square tests were used to evaluate for significant differences in uniformity by month of the year. Monthly trends were also stratified by gender for analysis. A p-value of <0.05 was taken as statistically significant.

RESULTS:

There were 4,595,616 patient visits in the database with 21,494 (null%) having an ICD-based diagnosis of acute myocardial infarction. In 2005, the most frequent ICD diagnosis code was null. In 2016, the most frequent ICD diagnosis was null null (Table S1, Fig. S1).

Fig 1. STROBE diagram detailing patient selection for acute myocardial infarction study.

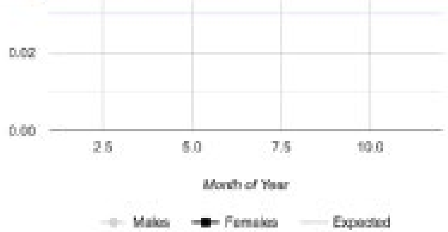
null


OUTPUT
healthlabs-4.appspot.com
Template
Workers
Links
User
Refs
Statics
Compose
DAG
Temporal Trends and Seasonal Patterns in the Incidence of Acute Myocardial Infarction from 2005-2016
Abstract
Introduction
Methods
Temporal Trends and Seasonal Patterns in the Incidence of Opiate Overdose from 2005-2016
Abstract
Introduction
Methods

Results onc_trend_myocardial | New Flight

Not Secure | healthlabs-4.appspot.com/view_results.html?w...

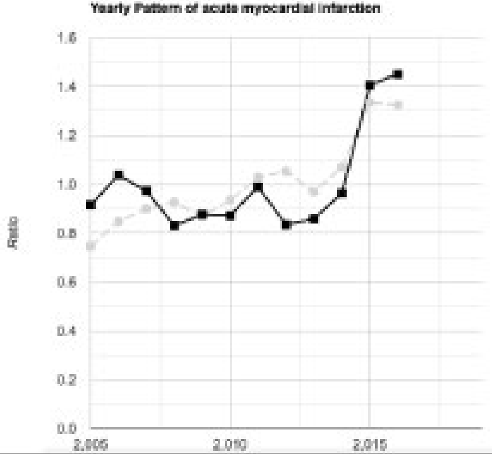
OUTPUT

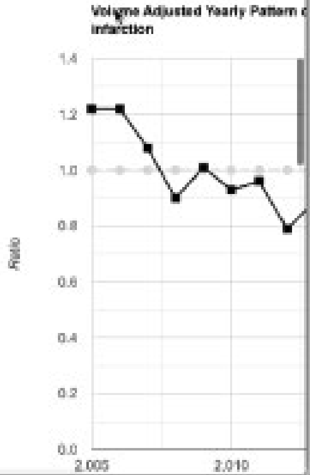


Year

Gross visits peaked in null (n=2596) and were lowest in null (n=1487), although when these figures were adjusted by ED visit numbers for each year, the proportion of ED visits that were associated with acute myocardial infarction generally decreased over time (Figures 6a and b)

***Figures 6a & b.* Gross and volume-adjusted trends in visits for AMI and visits for all causes across the time period 2014-2020.**



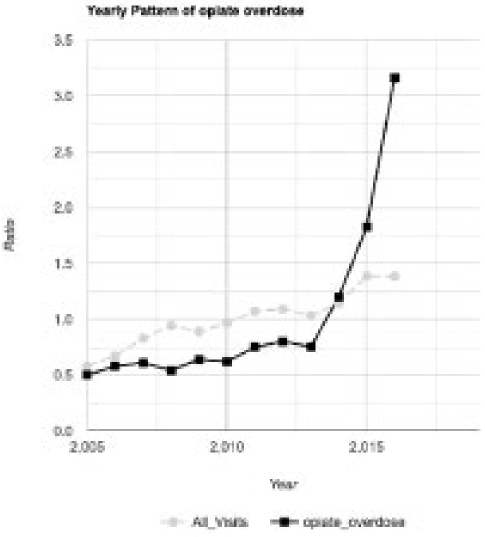


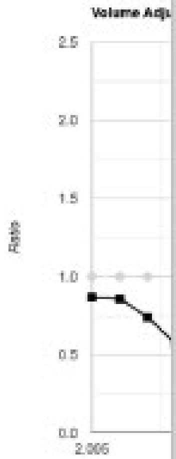
Not Secure | healthlabs-4.appspot.com/view_results...

Year

Gross visits peaked in 2016 (n=null) and were lowest in 2005 (n=290), although when these figures were adjusted by ED visit numbers for each year, the proportion of ED visits that were associated with opiate overdose generally decreased over time (Figures 6a and b)

***Figures 6a & b.* Gross and volume-adjusted trends in visits for OA and visits for all causes across the time period 2014-2020.**





There were definite temporal patterns in the incidence of ED visits for OA. Differences by hour of day were significant. Monthly differences were significant (p <.0001). There was no significant gender

ANTIBIOGRAM ENGINES

Federal Action Needed to Address Antibiotic Resistance in Older Americans

New research shows significant mortality, health care costs in this at-risk population

ISSUE BRIEF October 7, 2021 Topics: Antibiotics & U.S. Policy Projects: Antibiotic Resistance Tags: Superbugs Read time: 9 min

Singapore

Antibiotic resistance: UTI, among top 10 causes of death in Singapore, trickier to treat

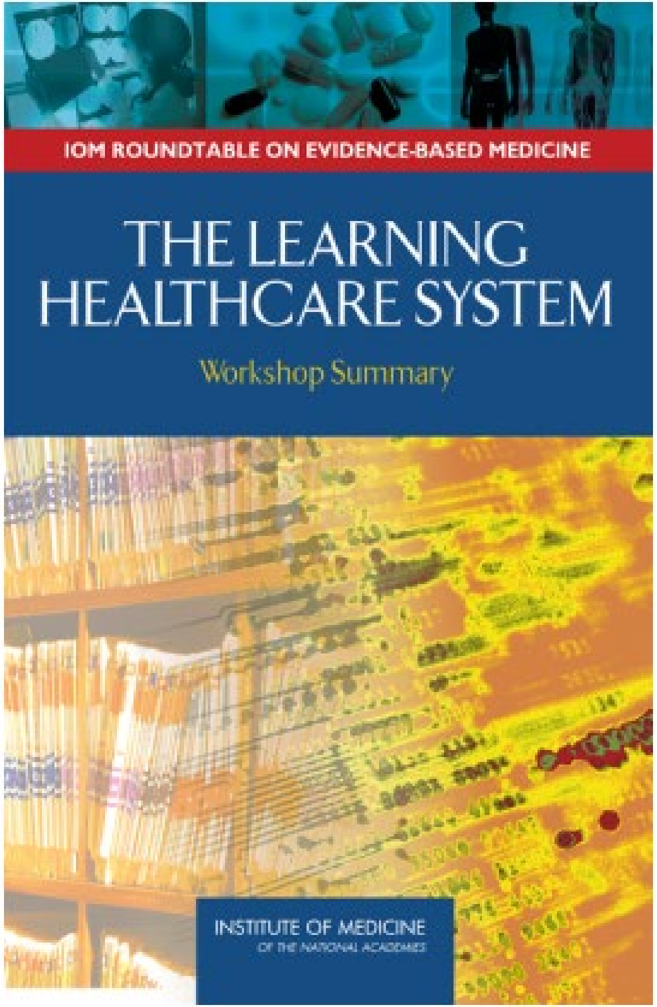
Superbugs in Seniors Led to Nearly 12,000 Deaths in a Single Year

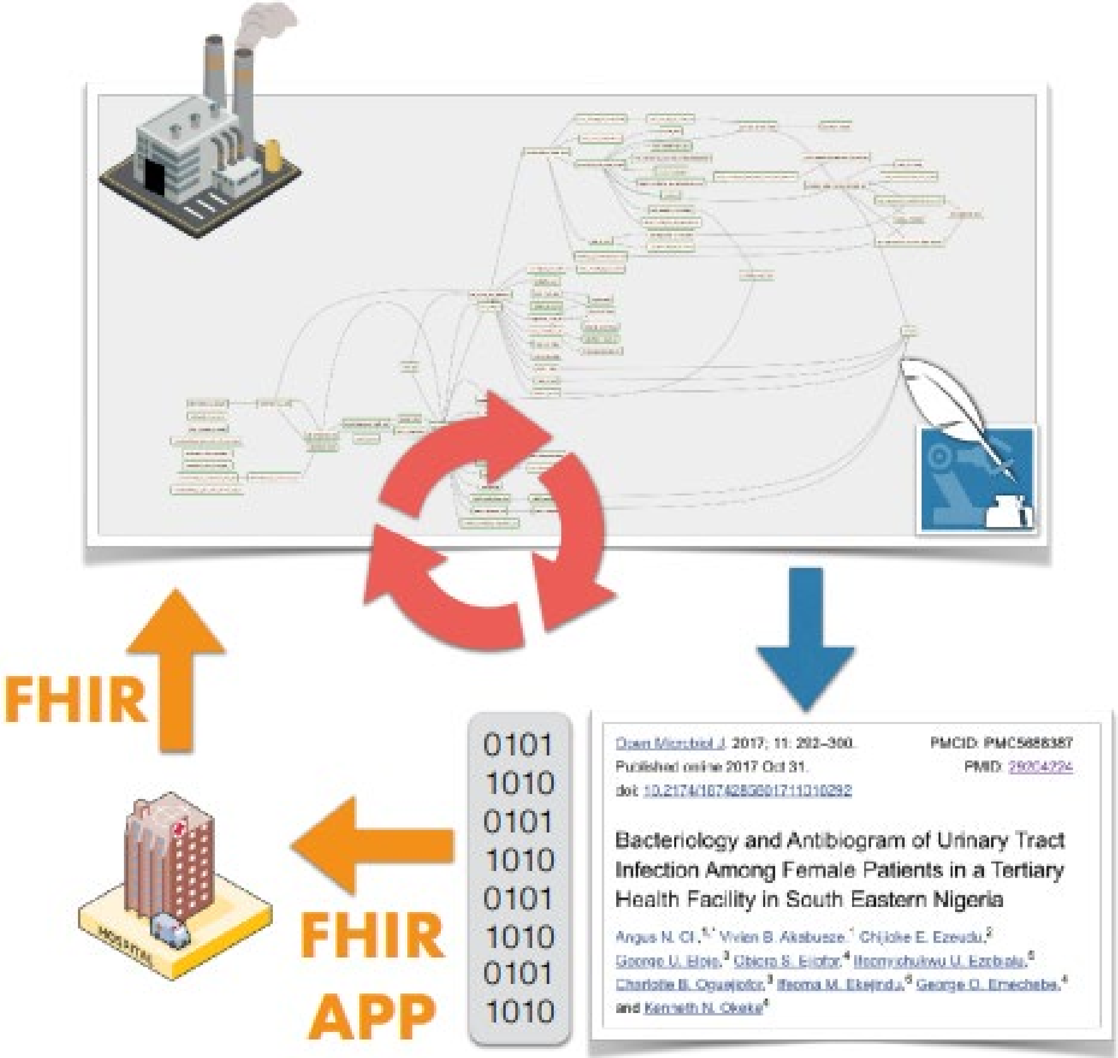
— Antibiotic-resistant infections in Medicare-age patients cost health system $2 billion in 2017

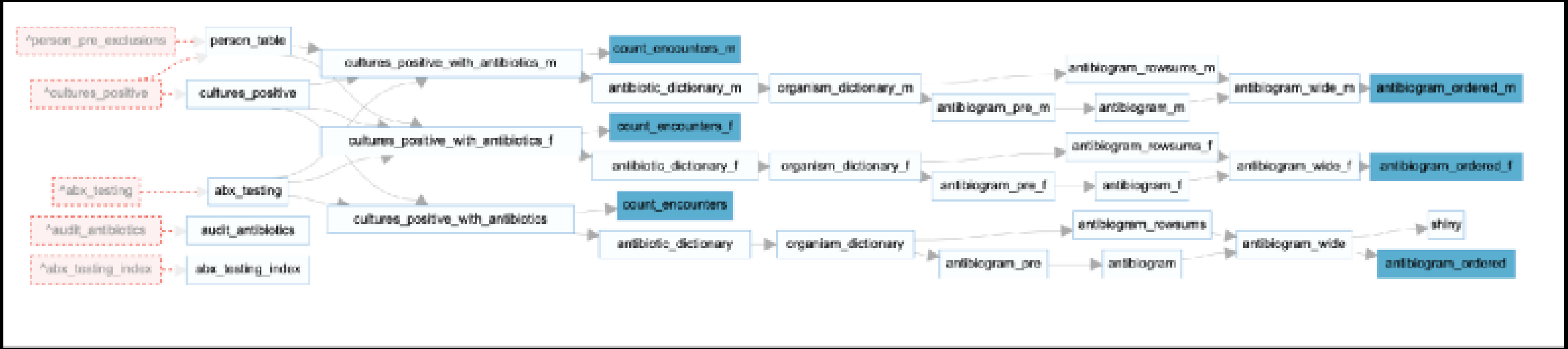
by Ryan Basen, Enterprise & Investigative Writer, MedPage Today
October 7, 2021

Europe's Antibiotic Investments Stagnate As Resistance Crisis Looms

BY JONATHAN SMITH
21/09/2021 - 7 MINUTES


IOM ROUNDTABLE ON EVIDENCE-BASED MEDICINE
THE LEARNING HEALTHCARE SYSTEM
Workshop Summary
INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES


FHIR
FHIR
APP
0101
1010
0101
1010
0101
1010
0101
1010
Published online 2017 Oct 31.
Bacteriology and Antibiogram of Urinary Tract Infection Among Female Patients in a Tertiary Health Facility in South Eastern Nigeria


^person_pre_exclusions
person_table
^cultures_positive
cultures_positive
^abx_testing
abx_testing
^audit_antibiotics
audit_antibiotics
^abx_testing_index
abx_testing_index
cultures_positive_with_antibiotics_m
cultures_positive_with_antibiotics_f
cultures_positive_with_antibiotics
count_encounters_m
count_encounters_f
count_encounters
antibiotic_dictionary_m
organism_dictionary_m
antibiotic_dictionary_f
organism_dictionary_f
antibiotic_dictionary
organism_dictionary
antibiogram_rowsums_m
antibiogram_pre_m
antibiogram_m
antibiogram_wide_m
antibiogram_ordered_m
antibiogram_rowsums_f
antibiogram_pre_f
antibiogram_f
antibiogram_wide_f
antibiogram_ordered_f
antibiogram_rowsums
antibiogram_pre
antibiogram
antibiogram_wide
shiny
antibiogram_ordered

FACTORY

Template | Workers | Links | User | Refs | Statics | Compose | DAG

Trends in susceptibility to antibiotics of Urinary Tract Infections

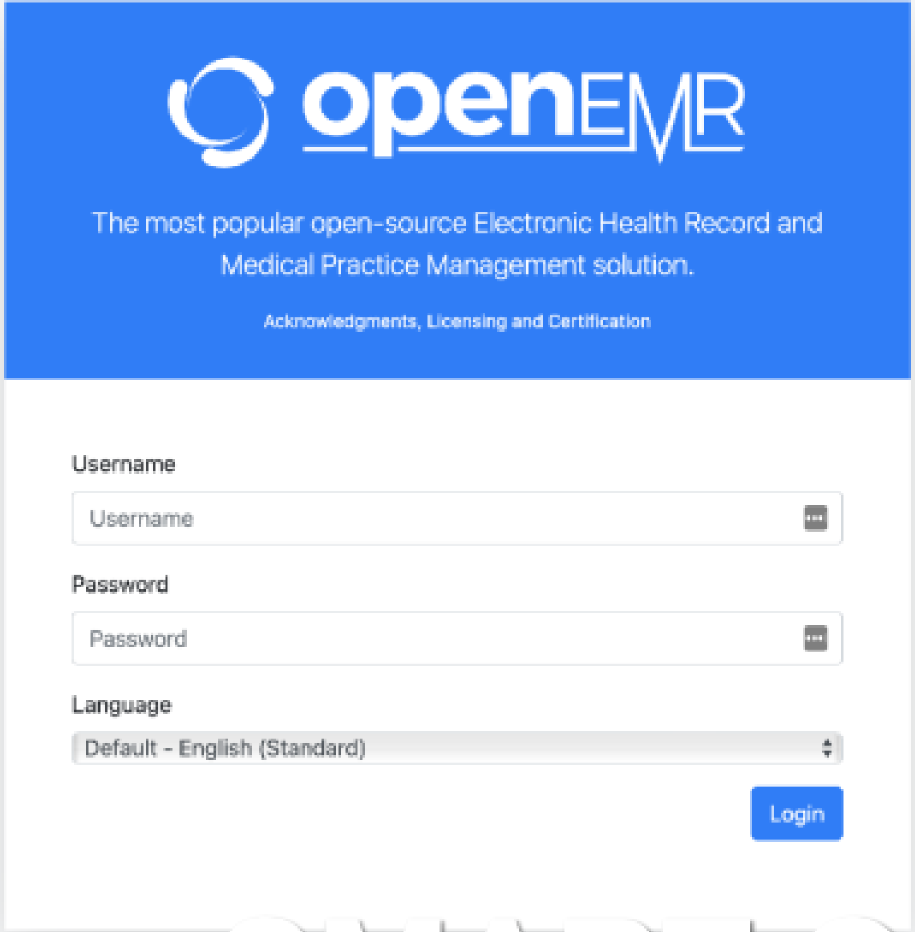
Number of encounters included: 24442 (7555 **male,** 16882 **female)**

Results

Table 1: Sensitivity testing frequency and results.
'n' refers to number of patient encounters for which tests were performed. Bracketed numbers include number of tests performed for each organism and antibiotic combination. Non-bracketed numbers refer to the percentage of test results that showed antibiotic sensitivity. 'R' indicates all results resistant.

ORGANISM	n	Amikacin	Amoxicillin	Amoxicillin_clavulanate	Ampicillin	Ampicillin_
Escherichia Coli	395661	99.59 (20681)	0	70.11 (20632)	40.09 (20680)	66.67 (6)
Klebsiella pneumoniae	151427	96.85 (8012)	0	65.87 (7993)	0.02 (8007)	66.67 (3)
Enterococcus faecalis	72272	0	0	0	96.49 (8455)	0
Proteus mirabilis	60025	97.52 (3507)	0	92.33 (3495)	71.91 (3510)	0
Pseudomonas aeruginosa	41904	97.66 (3718)	0	19.23 (26)	20.83 (24)	0
Enterococcus cloaceae	20353	99.91 (1131)	0	0.09 (1131)	10 (10)	0
Klebsiella oxytoca	12492	100 (655)	0	83.66 (655)	[illegible]	[illegible]
Enterococcus aerogenes	8624	99.61 (519)	0	R	R	[illegible]

OUTPUT



SMART ON FHIR

Antibiogram

Percent (%) Sensitive to an Antibiotic for a Male Patient

Number of cases in parenthesis.

Click here to view full table

ORGANISM	n	Ampicillin	Cefepime	Ciprofloxacin	Levofloxacin	Nitrofurantoin	Trimethoprim_sulfamethoxazole
Escherichia Coli	370875	27.64 (19550)	68.24 (19486)	41.09 (19546)	44.02 (10999)	94.07 (19431)	52.49 (19542)
Klebsiella pneumoniae	322068	- (17022)	41.93 (16830)	48.23 (17007)	56.72 (9655)	22.3 (16671)	44.45 (17013)
Pseudomonas aeruginosa	155397	31.25 (48)	73.78 (13925)	54.42 (13951)	48.01 (6346)	22.06 (68)	51.47 (68)
Enterococcus faecalis	150313	97.71 (17613)	0	0	48.57 (10521)	99.51 (16790)	R
Proteus mirabilis	117224	60.55 (6783)	93.59 (6783)	37.02 (6777)	37.26 (5011)	R	63.67 (6780)
Enterococcus cloaceae	40858	R	84.96 (2300)	84.64 (2304)	92.18 (844)	28.49 (2292)	63.5 (2304)
Proteus spp.	33534	R	95.79 (2020)	29.75 (2020)	26.05 (1359)	R	72.27 (1976)
Klebsiella oxytoca	29678	R	86.91 (1597)	86.57 (1593)	98.51 (470)	73.89 (1593)	85.6 (1597)
Enterococcus spp.	20900	48.67 (1730)	39.66 (585)	25.72 (587)	16.7 (1168)	83.23 (829)	41.54 (260)
MRSA	16853	R	0	0	100 (1)	100 (1436)	92.58 (1537)

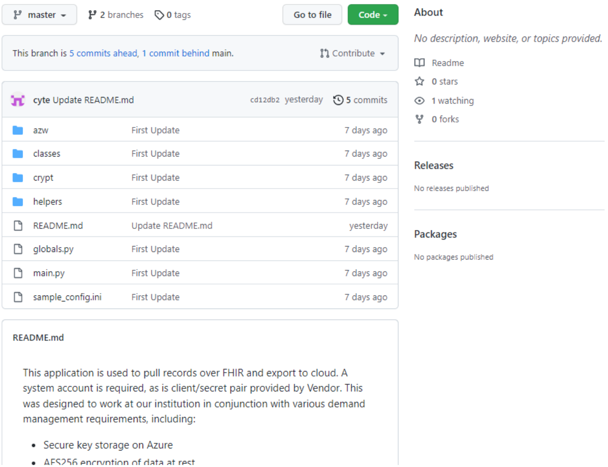
SMART ON FHIR

Dissemination and Links

Kristen.E.Miller@medstar.net

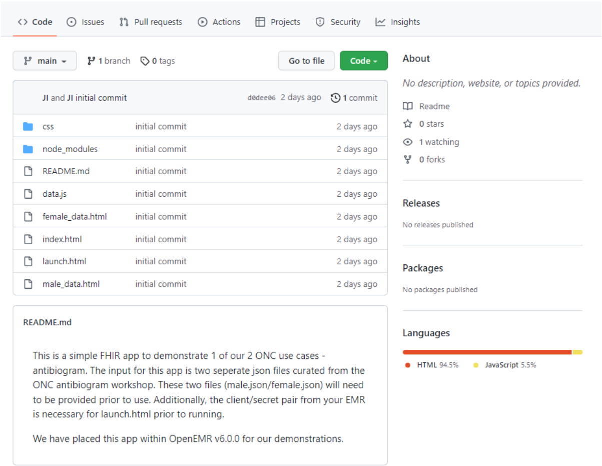
Trend Engine

https://github.com/cyte/hlfhir_public/tree/master/



Antibiogram App

https://github.com/cyte/oemr_fhir_antibiogram



SMART®
Cumulus

Instrumenting the healthcare system for population scale studies

https://docs.smarthealthit.org/cumulus/

Andy McMurry, PhD

Research Scientist, Boston Children's Hospital

Lecturer in Pediatrics, Harvard Medical School

Boston Children's Hospital

chip Computational Health Informatics Program

SMART®

Outline

- **Multi-solving**
  - Public health, quality, clinical research
- **Why now: what changed in 2023?**
  - Bulk FHIR in the cloud meets the AI race
- **How to engender participation?**
  - Healthcare site autonomy
  - Patient privacy (DEID)
  - Share population statistics
- **SMART Cumulus**
  - Open source architecture
  - Case definitions (study criteria)
  - Computable phenotypes (AI/NLP derived)
- **Accomplishments and Demo**
  - Bulk-FHIR standards process
  - Public health dyads pilot in 5 USA cities, 4 study areas
    - Mental Health, COVID19, HTN, OUD

Why now: what changed in 2023?

(1) **21st Century Cures Act** mandated **Bulk FHIR** API access
Jan 1, 2023 for **every EHR** in USA with "**no special effort**"

(2) **Cloud** adoption in healthcare:
write once, deploy in many healthcare sites.
"Turn-key" deployment of services and containers

(3) **AI/NLP** innovation speed
(ChatGPT, LLM, GNN, neural networks, ...)

Boston Children's Hospital
chip Computational Health Informatics Program
SMART®

A SMART Evolution

2009
NEJM: *No Small Change for the Health Information Economy* – introduces the API

White Paper: "... Fostering Development of an "iPhone-like" Platform..."

iTDotHealth meeting with government, vendors, academia

2010
$15M SHARP grant to the SMART Team

Draft SMART API released (RDF data models)

Cardiac Risk app becomes first SMART app

2011
SMART Apps Contest on Challenge.gov, promoted by White House. First health-related apps challenge

SMART Sandbox launched

2012
BP Centiles: First SMART App in production at Boston Children's Hospital

2013
Release of SMART on FHIR Genomics

SMART team joins FHIR development effort

2014
Release of SMART on FHIR specification

Launch of SMART Advisory Committee with diverse stakeholders committed to SMART on FHIR ecosystem

JASON Task Force report recommends a public API for healthcare

Argonaut project launched to implant SMART in EHRs

2015
2015 EHR certification and meaningful use 3 final rules require patient access via API

Launch of SMART App Gallery

Launch of SMART CDS Hooks, a decision support specification

2016
Precision Medicine Initiative Sync for Science uses SMART API

21st Century Cures Act makes APIs a requirement for certified HIT, incorporating language from the SMART team

Cerner launches SMART on FHIR Developers Sandbox

2017
Allscripts and EPIC launch SMART on FHIR developers sandboxes

Project Argonaut commits to CDS-Hooks Implementation

SMART Team, ONC & HL7 launch the SMART FLAT FHIR bulk data project

Carin Alliance focuses on implementing

2018
Apple adds SMART support to its Health App for patient access to medical data

CMS launches Blue Button 2.0 using SMART for API access to Medicare claims

Dept of Veterans Affairs to incorporate SMART on FHIR support into next-generation EHR platform

CMS pilots SMART bulk data FLAT FHIR export

SMART on FHIR published as an HL7 standard

2019
CMS commits to SMART Bulk Data specs

Microsoft launches SMART on FHIR API in its Azure product

Final rule from ONC specifies SMART as the universal apps API to implement 21st Century Cures

3 Cloud vendors at White House commit to SMART and FHIR open APIs

Veterans accessing health records using Apple's SMART on FHIR Health App

2020
SMART Markers Framework for Patient Generated Data

2021
Information blocking rules take effect

Epic launches initial version of SMART/HL7 Bulk FHIR Access API

SMART Health Cards established as a verified credential standard

VCI initiative establishes SMART Health Cards as an industry standard for vaccine credentials

The Commons Project releases the SMART Health Card verifier.

2022
Support for SMART on FHIR and Bulk FHIR APIs required in all certified HIT

2023
Support for full export of Electronic Health Information

SMART®

How to engender participation at healthcare system scale?

- **Healthcare provider autonomy**
  Each site remains in control of their:
  patient data, systems, policies, procedures

- **Privacy preserving data sharing**
  HIPAA De-identification
  Potential patient benefit

- **Share population health statistics**
  Broad agreement to share patient counts
  Population health dashboards

Federalist principles for healthcare data networks

Kenneth D Mandl & Isaac S Kohane

Nature Biotechnology **33**, 360–363 (2015) | Cite this article

Journal of the American Medical Informatics Association Volume 14 Number 4 July / August 2007

Model Formulation ■

A Self-scaling, Distributed Information Architecture for Public Health, Research, and Clinical Care

Andrew J. McMurry, Clint A. Gilbert, Ben Y. Reis, PhD, Henry C. Chueh, MD, MS, Isaac S. Kohane, MD, PhD, Kenneth D. Mandl, MD, MPH

Boston Children's Hospital

Computational Health Informatics Program

SMART®

How does Cumulus work?

Installation → Initial Queries → NLP → Data Deidentification → Aggregate Counts → Visualization & Analysis

Installation

The Cumulus client is installed behind the firewall of a healthcare site, health information exchange (HIE), etc.

Initial Queries

A query is run, pulling electronic health data into a secure server behind the organization's firewall.

NLP

AI-based natural language processing (NLP) is run on the clinical notes to extract clinical information from the text.

Data Deidentification

Data from the EHR (both discrete data from things like lab values and the NLP results) are de-identified.

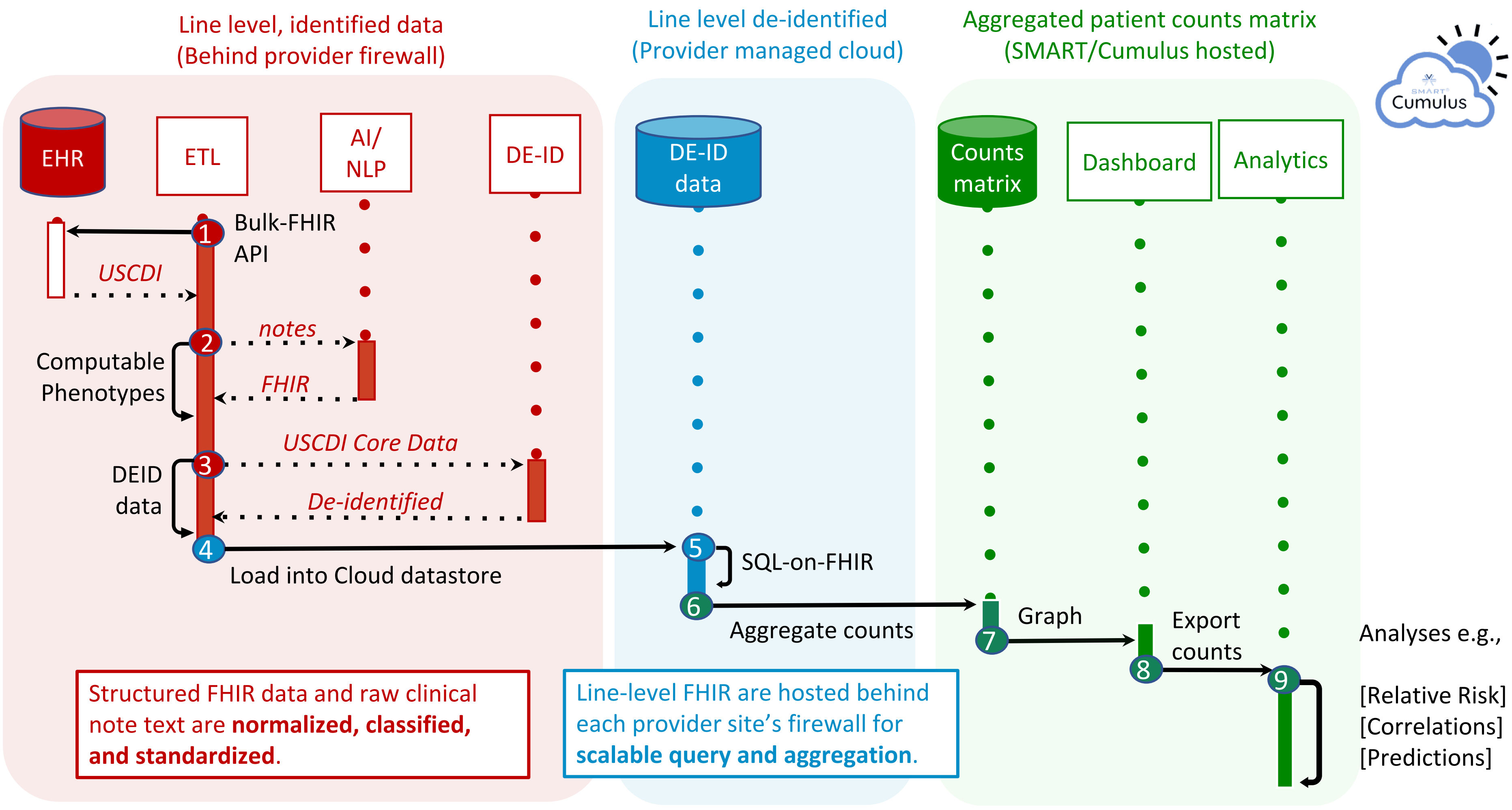
Aggregate Counts

SQL (Structured Query Language) is used to generate counts. Note – this still takes place under the healthcare site's control.

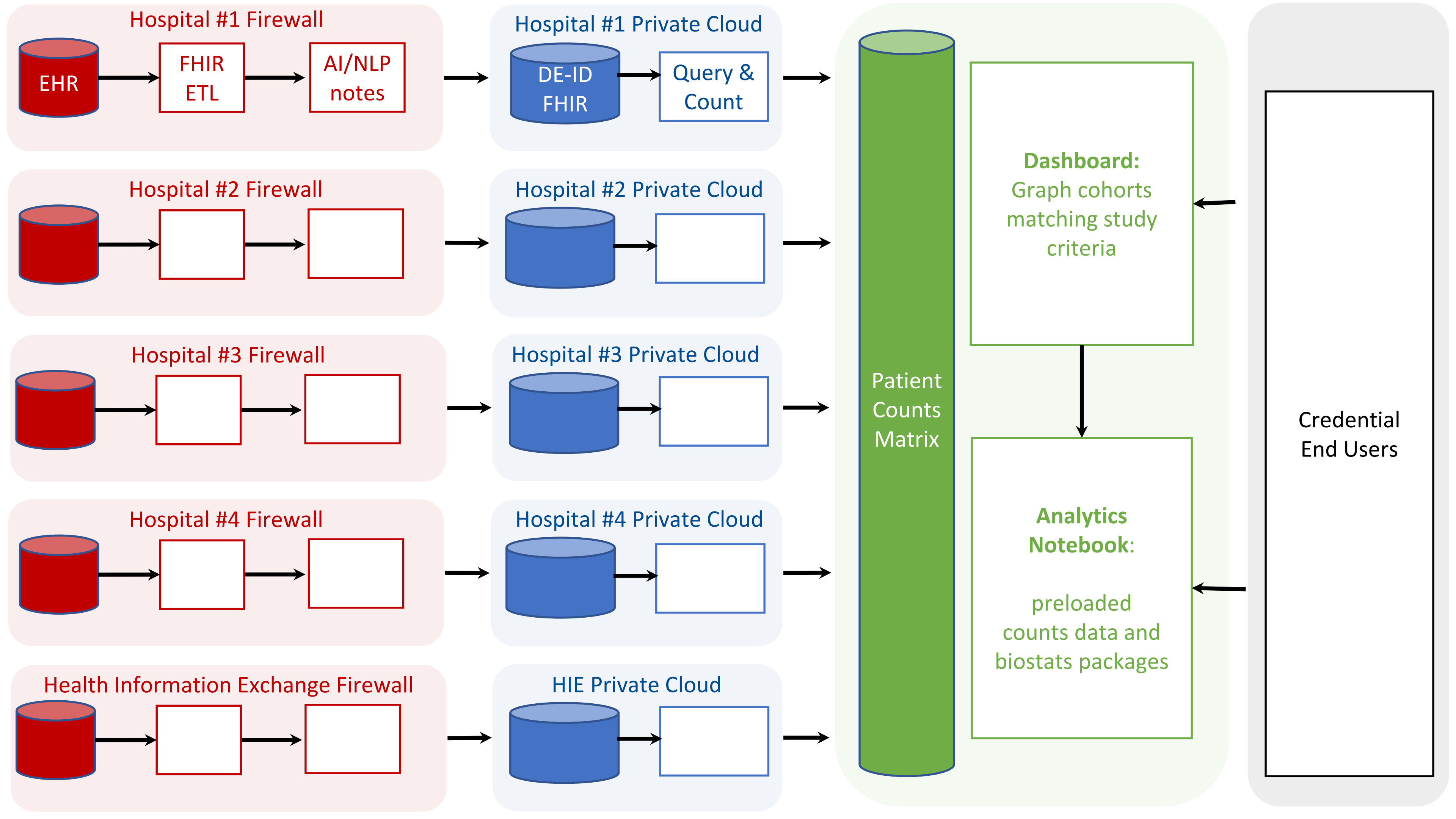
Visualization & Analysis

A web-based Cumulus dashboard is available for public health use (e.g., data visualization and analytics).

Counts data leaves the healthcare site to an external Cumulus server



SMART Cumulus Architecture


Hospital #1 Firewall
EHR
FHIR ETL
AI/NLP notes
Hospital #1 Private Cloud
DE-ID FHIR
Query & Count
Hospital #2 Firewall
Hospital #2 Private Cloud
Hospital #3 Firewall
Hospital #3 Private Cloud
Hospital #4 Firewall
Hospital #4 Private Cloud
Health Information Exchange Firewall
HIE Private Cloud
Patient Counts Matrix
Dashboard: Graph cohorts matching study criteria
Analytics Notebook: preloaded counts data and biostats packages
Credential End Users

Cumulus Open Source Repos

- **ETL**
  - Extract **Bulk-FHIR** data from local EHR
  - AI/NLP **computable phenotypes**
  - DEID remove PHI/PII
  - Load into cloud datastore

- **Study Library**
  - **SQL-on-FHIR** simplifies query and analysis
  - **Case definitions**, study criteria, “counts” query
  - Support existing value sets (VSAC)
  - Propensity score matching (PSM)

- **Aggregator**
  - Aggregate **patient counts matrix** from multiple sites

- **Dashboard**
  - **Graph**, analyze, filter, and compare patient populations

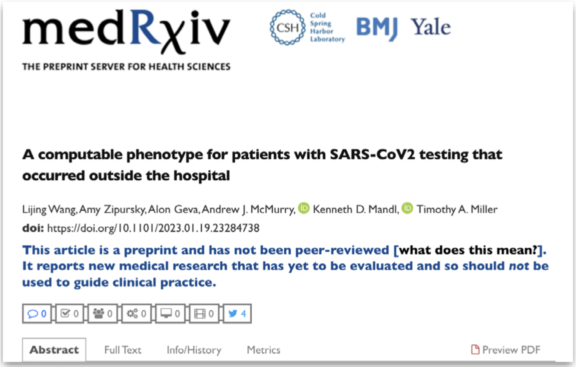
SMART® Cumulus

https://docs.smarthealthit.org/cumulus/

AI/NLP "Computable Phenotypes"

EHR data without computation do not yield precise diagnoses, risk factors, endpoints, etc.

Computable case definitions are often based on multiple elements which can include NLP of free text


medRχiv
THE PREPRINT SERVER FOR HEALTH SCIENCES
CSH Cold Spring Harbor Laboratory BMJ Yale

A computable phenotype for patients with SARS-CoV2 testing that occurred outside the hospital

Lijing Wang, Amy Zipursky, Alon Geva, Andrew J. McMurry, Kenneth D. Mandl, Timothy A. Miller
doi: https://doi.org/10.1101/2023.01.19.23284738

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

0 0 0 0 0 0 4

Abstract Full Text Info/History Metrics Preview PDF

In an expert validation, the classifier correctly identified 90.8% (79/87) as COVID-19 positive and 97.8% (91/93) as not SARS-CoV2 positive. The classifier identified an additional 960 positive cases that did not have SARS-CoV2 lab tests in hospital, and only 177 of those cases had the ICD-10 code for COVID-19.

Boston Children's Hospital
chip Computational Health Informatics Program
SMART®

AI/NLP “Computable Phenotypes”

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Cold Spring Harbor Laboratory BMJ Yale

The SMART Text2FHIR Pipeline

Timothy A. Miller, Andrew J. McMurry, James Jones, Daniel Gottlieb, Kenneth D. Mandl
doi: https://doi.org/10.1101/2023.03.21.23287499

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Cold Spring Harbor Laboratory BMJ Yale

Moving Biosurveillance Beyond Coded Data: AI for Symptom Detection from Physician Notes

Andrew McMurry, Amy R Zipursky, Alon Geva, Karen L Olson, James Jones, Vlad Ignatov, Timothy Miller, Kenneth D Mandl
doi: https://doi.org/10.1101/2023.09.24.23295960

Reason for Visit: Patient complains of fever, cough.

HPI: Patient is an 8-year-old female presenting today for worsening fever and cough. Patient denies sore throat, headache, fatigue.

PMH: Asthma since age 7

PSH: Tonsillectomy and Adenoidectomy, PDA closure, tympanostomy tube placement.

FHI: reviewed and non-contributory

SH: Lives at home with parents and older sister

Immunizations: Up to date

ROS:
Pertinent positives and negatives noted above in HPI. All other systems of a 10 system review are negative.

Home Medications:
Flovent, albuterol prn

Medications Prescribed This Visit:
acetaminophen (acetaminophen 160 mg/5 mL oral liquid), 131.2 mg = 4.1 mL, PO, Q4hr, PRN


```
{
  "id": "05cb5888-2c9a-41ad-b877-eaa48c018657",
  "extension": [ {
    "extension": [ {
      "url": "reference",
      "valueReference": {
        "reference": "DocumentReference/44ce9f27-f6ca-44f2-a79b-c88ba8abfe8b"
      }
    }, {
      "url": "offset",
      "valueInteger": 39
    }, {
      "url": "length",
      "valueInteger": 5
    } ],
    "url": "http://hl7.org/fhir/StructureDefinition/derivation-reference"
  } ],
  "modifierExtension": [ {
    "extension": [ {
      "url": "algorithm",
      "valueString": "ctakesclient"
    }, {
      "url": "version",
      "valueString":
"https://github.com/Machine-Learning-for-Medical-Language/ctakes-client-py/releases/tag/v2.0.0"
    } ],
    "url": "http://fhir-registry.smarthealthit.org/StructureDefinition/nlp-source"
  }, {
    "url": "http://fhir-registry.smarthealthit.org/StructureDefinition/nlp-polarity",
    "valueBoolean": true
  } ],
  "code": {
    "coding": [ {
      "code": "386661006",
      "system": "http://snomed.info/sct"
    }, {
      "code": "C0015967",
      "system": "http://terminology.hl7.org/CodeSystem/umls"
    }, {
      "code": "50177009",
      "system": "http://snomed.info/sct"
    }, {
      "code": "C0015967",
      "system": "http://terminology.hl7.org/CodeSystem/umls"
    }
    ],
    "text": "fever"
  },
  "encounter": {
    "reference": "Encounter/8aa2fe54-bfe1-4e3d-a645-bd9f4b5fec78"
  },
  "status": "preliminary",
  "subject": {
    "reference": "Patient/d1129431-29c9-488e-97a4-00c8c5a0a4b1"
  },
  "resourceType": "Observation"
},
```


Boston Children's Hospital

Computational Health Informatics Program

SMART®

AI/NLP “Computable Phenotypes”

Encounter Time	Include	Exclude
Prior to encounter	Chief complaint or patient reason for visit	Manifestations of chronic, pre-existing, or other known condition that is not COVID-19
Present ED encounter	Yes	Manifestations of chronic, pre-existing, or other known condition that is not COVID-19
Future encounter	Never	Always

ED Note Section	Include	Exclude
Chief Complaint	Symptom present	
HPI, ROS	Symptom present	
Physical Exam	Symptom present	
Vital Signs	Fever present	
History: medical, family, social		Always (e.g. brother had a cough)
Treatment		Always (e.g. albuterol PRN for cough)
Investigations		Always (e.g. opacity on chest x-ray)
Assessment and plan, course, evaluation	Symptom present	
Final diagnosis	Symptom present	
Discharge Instructions	Symptom present	

Symptom specific	Include	Exclude
New loss of taste or smell	Anosmia, loss of taste, loss of smell	Injury related loss of taste/smell
Congestion or runny nose	Rhinorrhea, congestion, or discharge. Nose is dripping, running, or stuffy.	
Cough	Tussive or post-tussive. Cough is unproductive, productive, dry, wet, or producing sputum.	Wheeze, crackles, croup
Diarrhea	Diarrhea or watery stool	Loose stool, bloody stool
Fatigue	Fatigue, tired, exhausted, weary, malaise, feeling generally unwell.	Looked ill
Fever or chills	Fever, pyrexia, chills, or temperature >= 100.4 °F [38 °C]	Afebrile, felt warm
Headache	HA/headache, migraine, cephalgia, head pain	Headache due to Injury
Muscle or body aches	Myalgias, myoneuralgia, muscles or body aches, soreness. Generalized aches and pains.	Localized pain, injury, ABD pain, lower back pain
Nausea or vomiting	Nausea, vomiting, emesis, throw up, queasy, regurgitated	Gastritis, gastroparesis
SOB or difficulty breathing	Dyspnea, breathing is short, difficult, increased, labored, or distressed	BiPAP, CPAP, or other oxygen assistive device
Sore throat	Sore throat, throat pain, pharyngeal pain, pharyngitis, odynophagia	Streptococcus, dysphagia, hoarseness, red throat

Patient Counts Matrix

SQL-on-FHIR:
Select x,y,z
fr
cr

<u>user study criteria</u>

With *powerset* as
(Select “study variables” where “study criteria”)

Cloud datastore

CSV File
Powerset Matrix
(weekly, monthly, or yearly)

SMART Cumulus

Dashboard

https://github.com/smart-on-fhir/cumulus-library

Patient Counts Matrix

Count frequency of

Exposure		Outcome: Yes	Outcome: No	Total
Exposure	Yes	A	B	
	No	C	D	
	Total			N

matrix

Contingency Tables
Count frequency of exposure/outcome

Conditional Probability, Entropy, Mutual Information
Probability (0-100%) of event co-occurence

Odds Ratio
Association of exposure and outcome

Relative Risk
Probability of an event occurring in the exposed group versus the probability of the event occurring in the non-exposed group

Comorbidities and other factors
Co-occurrence of conditions, manifestations, related variables

Chi-Square
Test hypothesis that two or more events are independent or correlated.

Bayesian Modeling
Naive bayes classifiers type models are “pre-computed” baseline models from conditional probabilities

Population Health
Inherently involves “counting” patients

What has been accomplished?

Technology standards - Team developed the FHIR Bulk Data API and reference implementations. Required by regulation. Working with EHR vendors to test and refine bulk data capability.

Beta software - Initial versions of open-source data pipeline, de-identification infrastructure, NLP engine, multi-site data aggregator, public health dashboard.

Pilot site network - Five sites integrating beta software with their EHR systems and turning on public health surveillance feeds for COVID and other study areas.

Pilot Dyads - Effectively launched Jan 2023

Boston Children's Hospital & Massachusetts Department of Public Health

Regenstrief Institute & Marion County Public Health Department

Rush University Medical Center & Chicago Department of Public Health

Washington University in St. Louis & St. Louis Department of Public Health

UC Davis & Yolo County & Sacramento County

Boston Children's Hospital Regenstrief Institute RUSH Washington University in St. Louis UCDAVIS UNIVERSITY OF CALIFORNIA

Cumulus Dashboard Examples

COVID19

NLP extract COVID19 PCR computable phenotype

NLP extract COVID19 symptoms computable phenotype

Hypertension

HTN Comorbidities

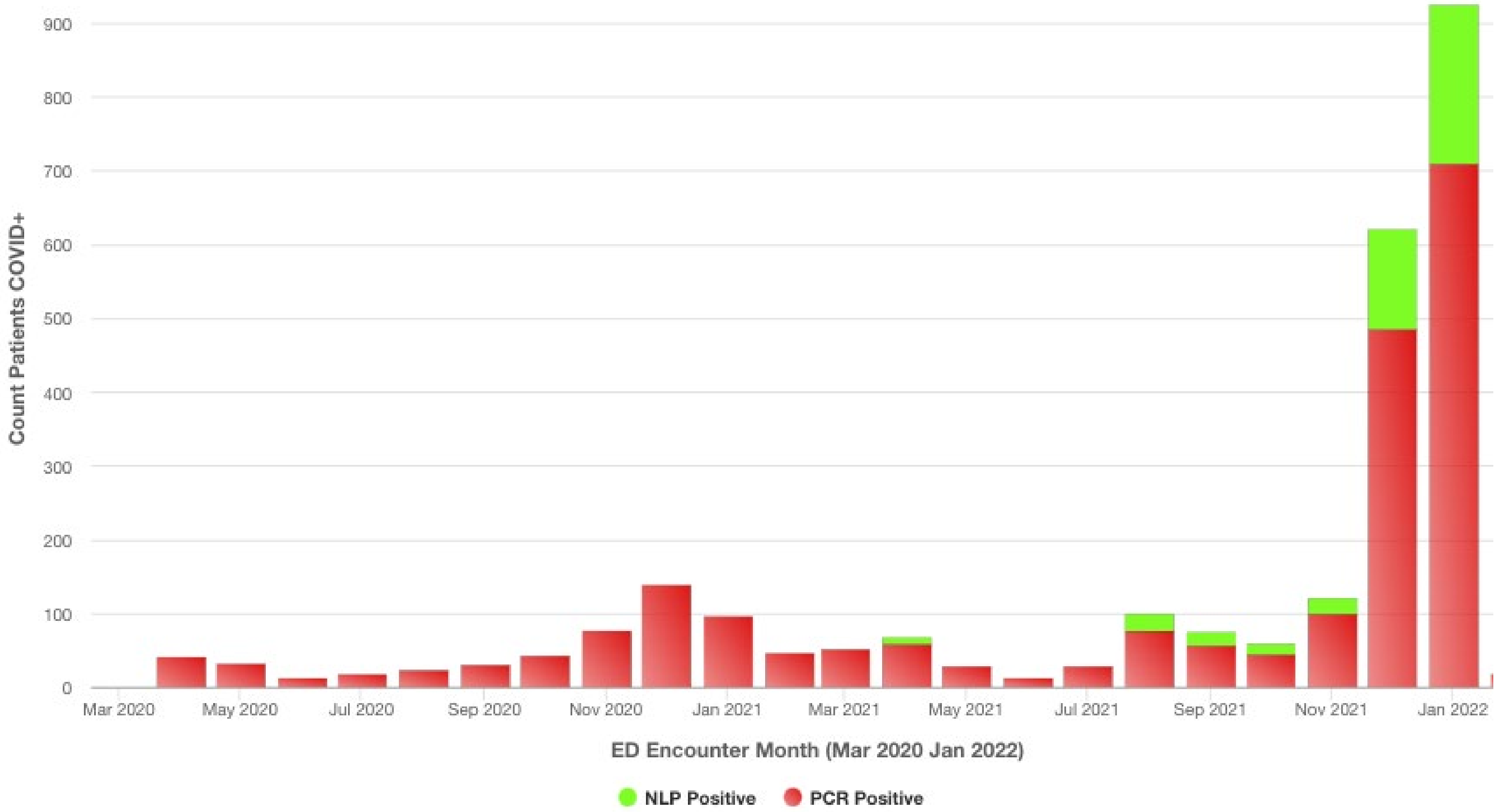
Suicidality

Prevalence in pediatric ED visits

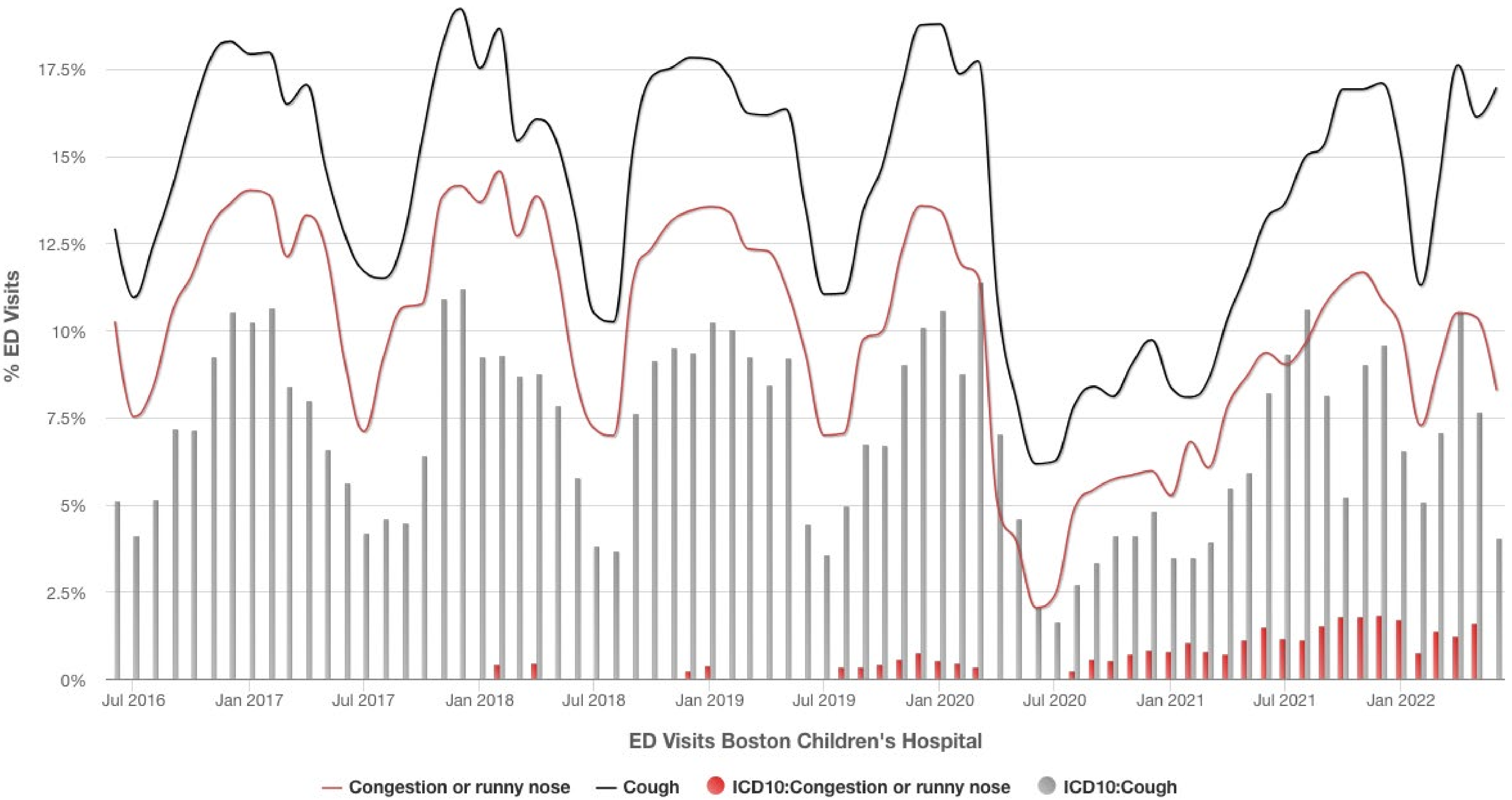
Opioid Overdose

Weekly Encounters

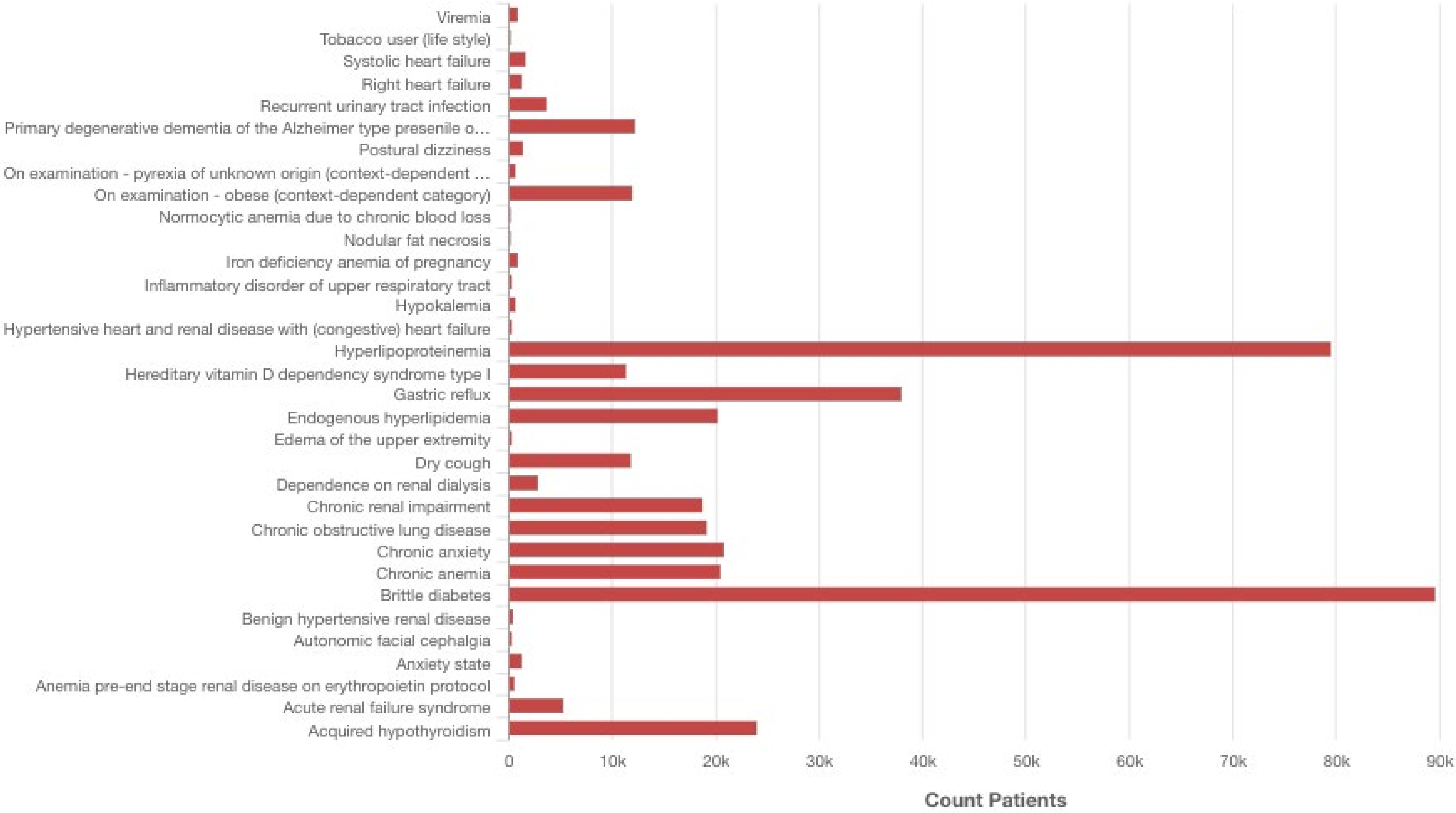
NLP extract **COVID19 PCR** status when PCR was performed outside the hospital



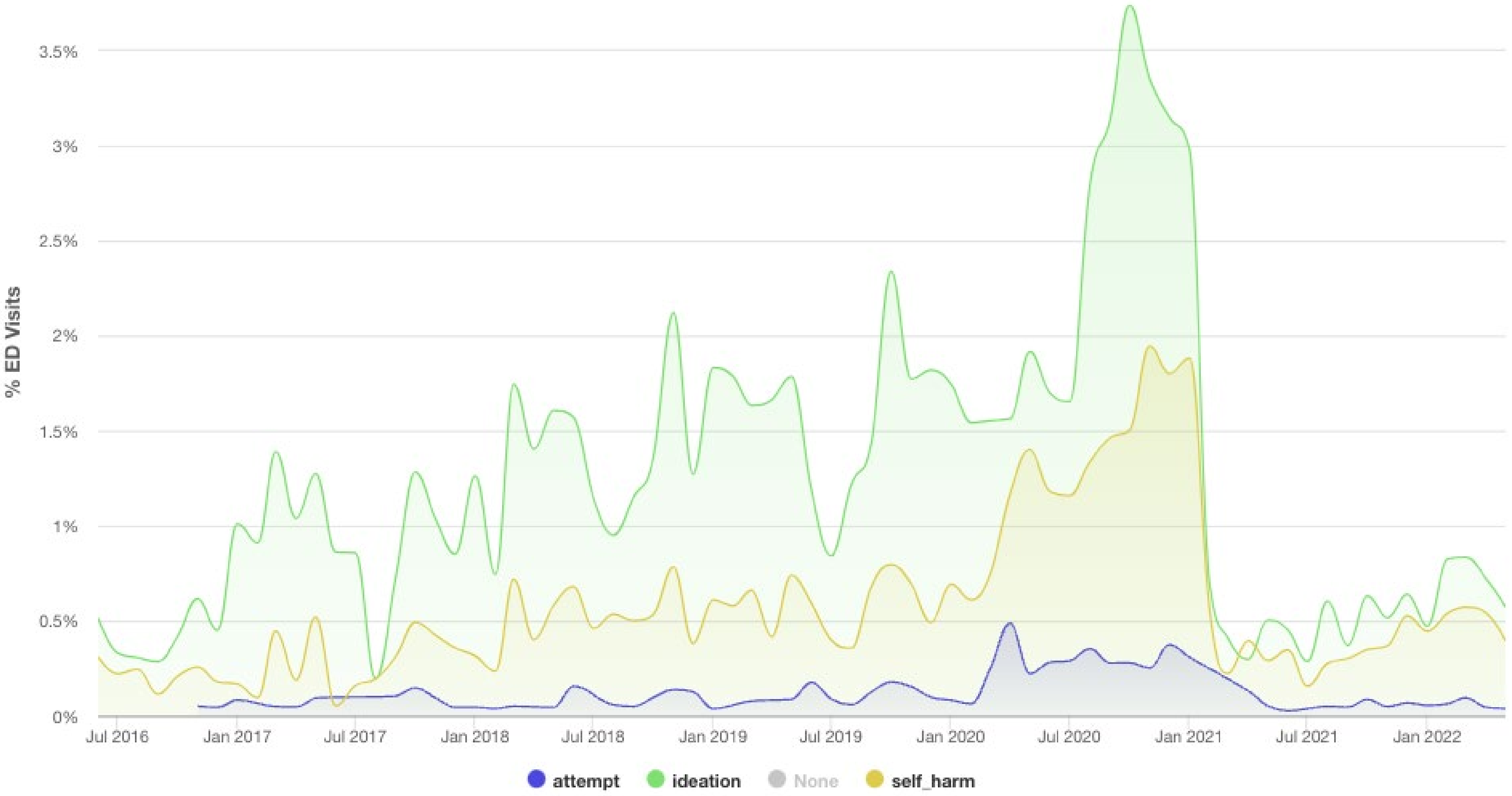
NLP [seasonal trend of COVID symptoms]



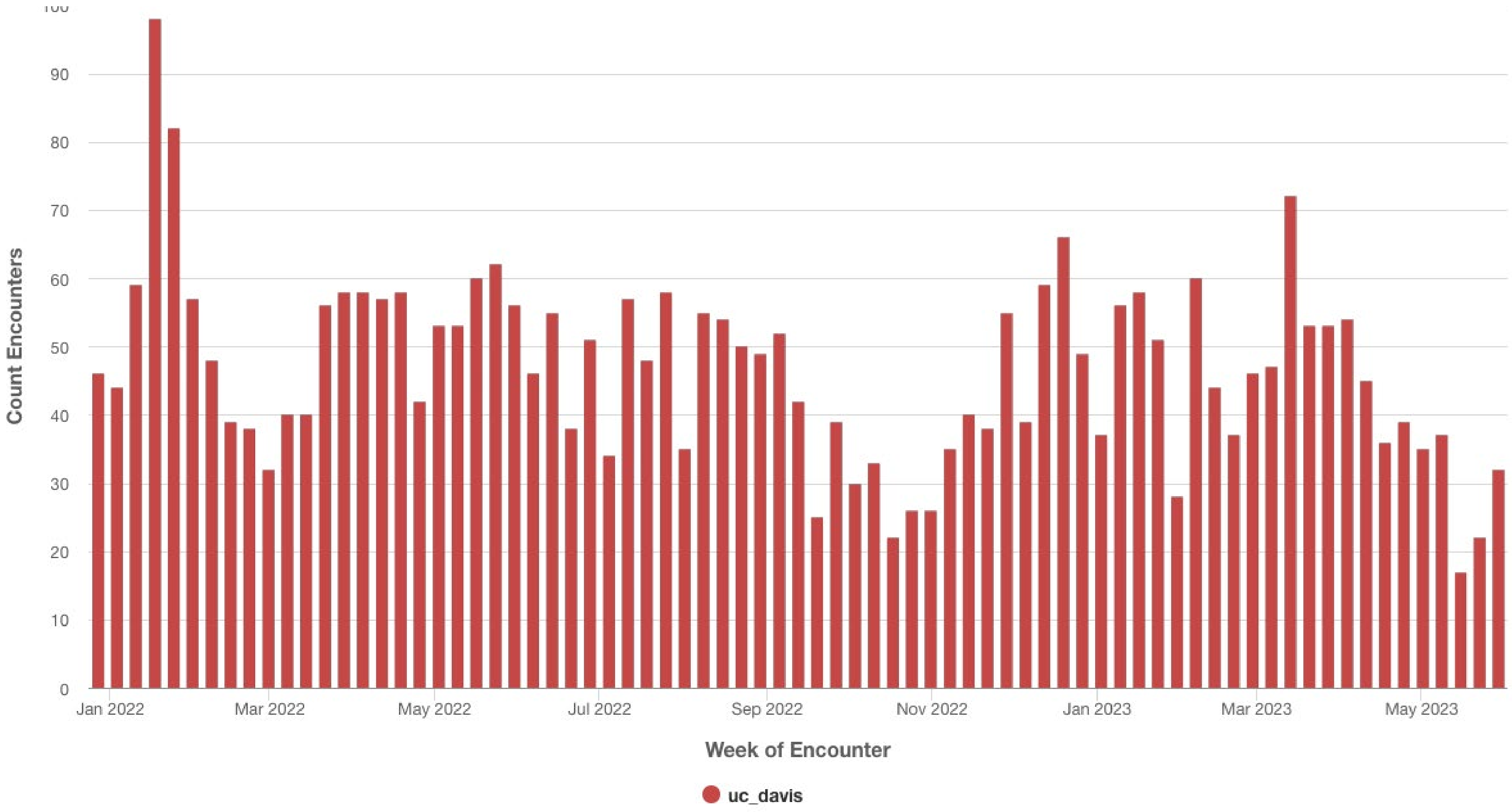
Top Hypertension comorbidities in COVID+ cases



Suicidality prevalence in ED visits



Weekly Encounters for Opioid Overdose



Summary

- **Multi-solving**
  - Public health, quality, clinical research

- **Why now: what changed in 2023?**
  - Bulk FHIR in the cloud meets the AI race

- **How to engender participation?**
  - Healthcare site autonomy
  - Patient privacy (DEID)
  - Share population statistics

- **SMART Cumulus**
  - Open source architecture
  - Case definitions (study criteria)
  - Computable phenotypes (AI/NLP derived)

- **Accomplishments and Demo**
  - Bulk-FHIR standards process
  - Public health dyads pilot in 5 USA cities, 4 study areas
    - Mental Health, COVID19, HTN, OUD

SMART®
Cumulus

Appendix

Intentionally Blank.

Semantic Interoperability of EHR Data Using the Layered Schema Architecture

ONC 2023 Annual Meeting
December 14, 2023

Burak Serdar
bserdar@cloudprivacylabs.com

Wilson Pace, MD – PI
Wilson.Pace@DARTNet.info

DARTNet Institute
Informing Practice Improving Care

Cloud Privacy Labs

The Problem

- Perhaps improving interoperability of EHR data
  BUT
- Clinical care and Researchers both rapidly adding more data types
- Cacophony of data for both groups is increasing not decreasing
- Traditional Translation and Loading approaches are not up to the challenge

DARTNet Institute
Informing Practice Improving Care

Cloud Privacy Labs

Use Cases

- **All of Us**
- 1 – 1.2 M people
- 20 year time span
- EHR data – multiple vendors
- Case Report Form data
- Social determinants data
- Genetic data
- Omics data
- Environmental data

- **Health Information Exchanges**
- Multiple data providers
  - Clinical organizations
  - Social services agencies
  - Community support organizations
  - Community wellness organizations
  - ? Justice systems
  - ? School health organizations
- All "speak" their own language
- Universal translator must ingest and store these data and provide it back to each user in a way they can read and interpret the information

DARTNet Institute
Informing Practice Improving Care

Schemas and Semantics

Cloud Privacy Labs

https://ohdsi.github.io/CommonDataModel/cdm54.html#PERSON

CDM Field	Datatype	Required	User Guide	ETL Conventions
person_id	integer	yes		
gender_concept_id	integer	yes	This field is meant to capture the biological sex at birth of the Person. This field should not be used to study gender identity issues.	Use the gender or sex value present in the data under the assumption that it is the biological sex at birth. If the source data captures gender identity it should be stored in the OBSERVATION table. Accepted gender concepts
gender_source_value	string	no	This field is used to store the biological sex of the person from the source data. It is not intended for use in standard analytics but for reference only.	Put the biological sex of the person as it appears in the source data.

Schema defines structure, not semantics

Semantics are human-readable

Layered Schemas

DARTNet Institute — Informing Practice Improving Care

Cloud Privacy Labs

Schema overlays capture metadata, context, and machine-readable semantics
Schema + a unique set of overlays for each distinct source
Data-agnostic, Standard-agnostic

Schema

CDM Field	Datatype	Required
person_id	integer	yes
gender_concept_id	integer	yes
gender_source_value	string	no

Overlays

CDM Field	ETL Conventions
gender_source_value	lookup in gender_dict

CDM Field	Privacy Level
gender_source_value	PII
person_id	PII

CDM Field	Confidence
ethnicity_source_value	.6

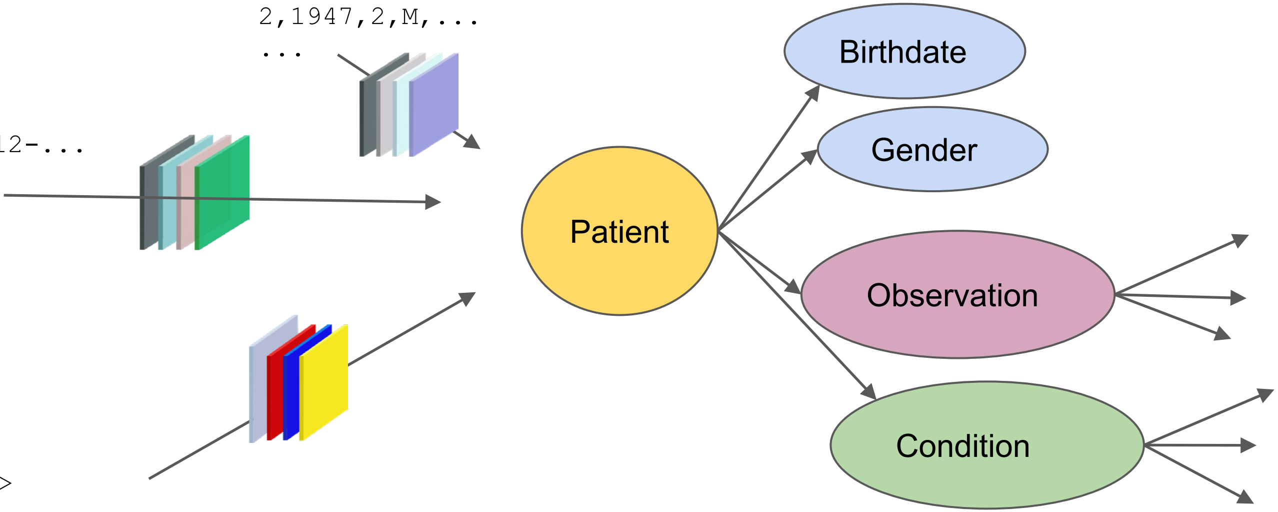
Person-Centric Graph Model

FHIR Patient (json)

```
],
"gender": "male",
"birthDate": "1974-12-25",
"_birthDate": {
  "extension": [
    {
      "valueDateTime": "1974-12-...
...
```

person.csv

```
person_id,year_of_birth,month_of_birth,gender,...
1,1974,12,M,...
2,1947,2,M,...
...
```



CCDA (xml)

```
<patient>
   <name>
      <given>...</given>
      <family>...</family>
   </name>
   <administrativeGenderCode code="M" codeSystem="2.16.840.1.113883.5.1"/>
   <birthTime value="19741225"/>
```

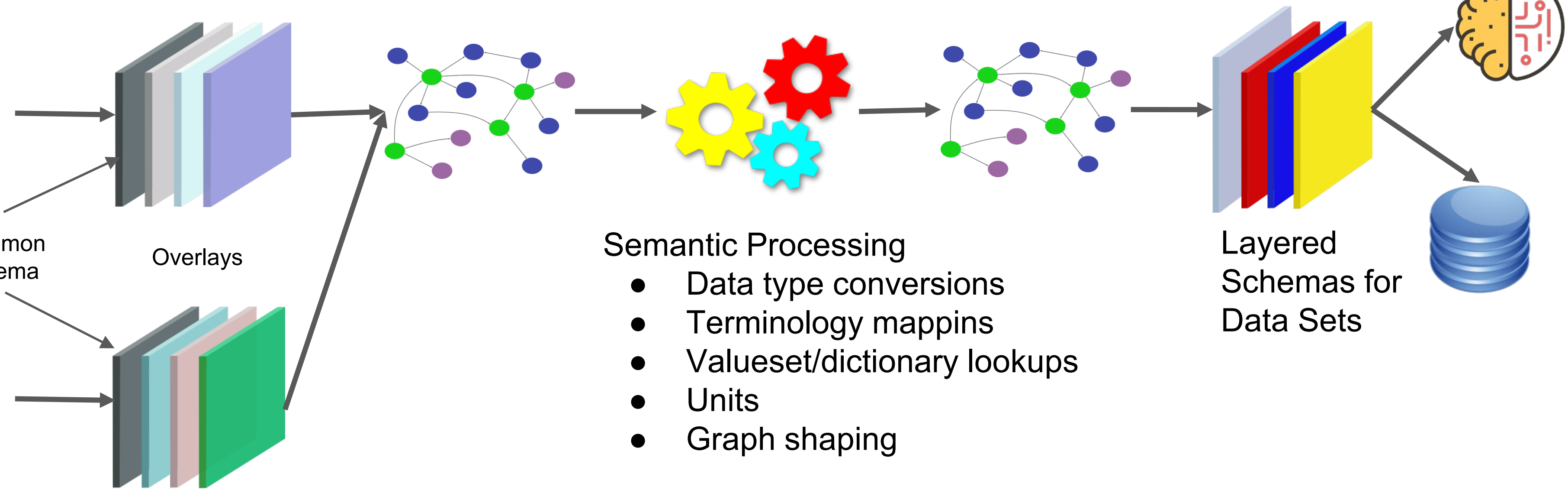
DARTNet Institute
Informing Practice Improving Care

Semantic Pipeline

Cloud Privacy Labs

Ingest, annotate, and harmonize data as a knowledge graph

Translate graph to research data sets



ONC LEAP - Vision

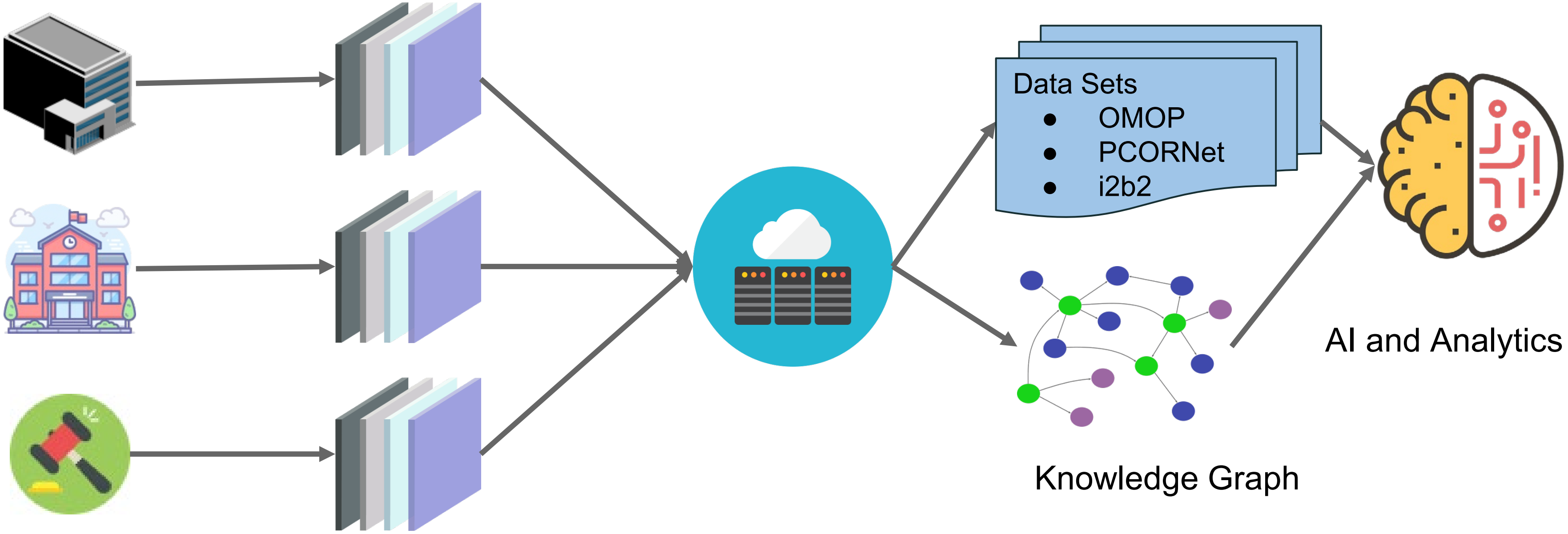
DARTNet Institute
Informing Practice Improving Care

Cloud Privacy Labs

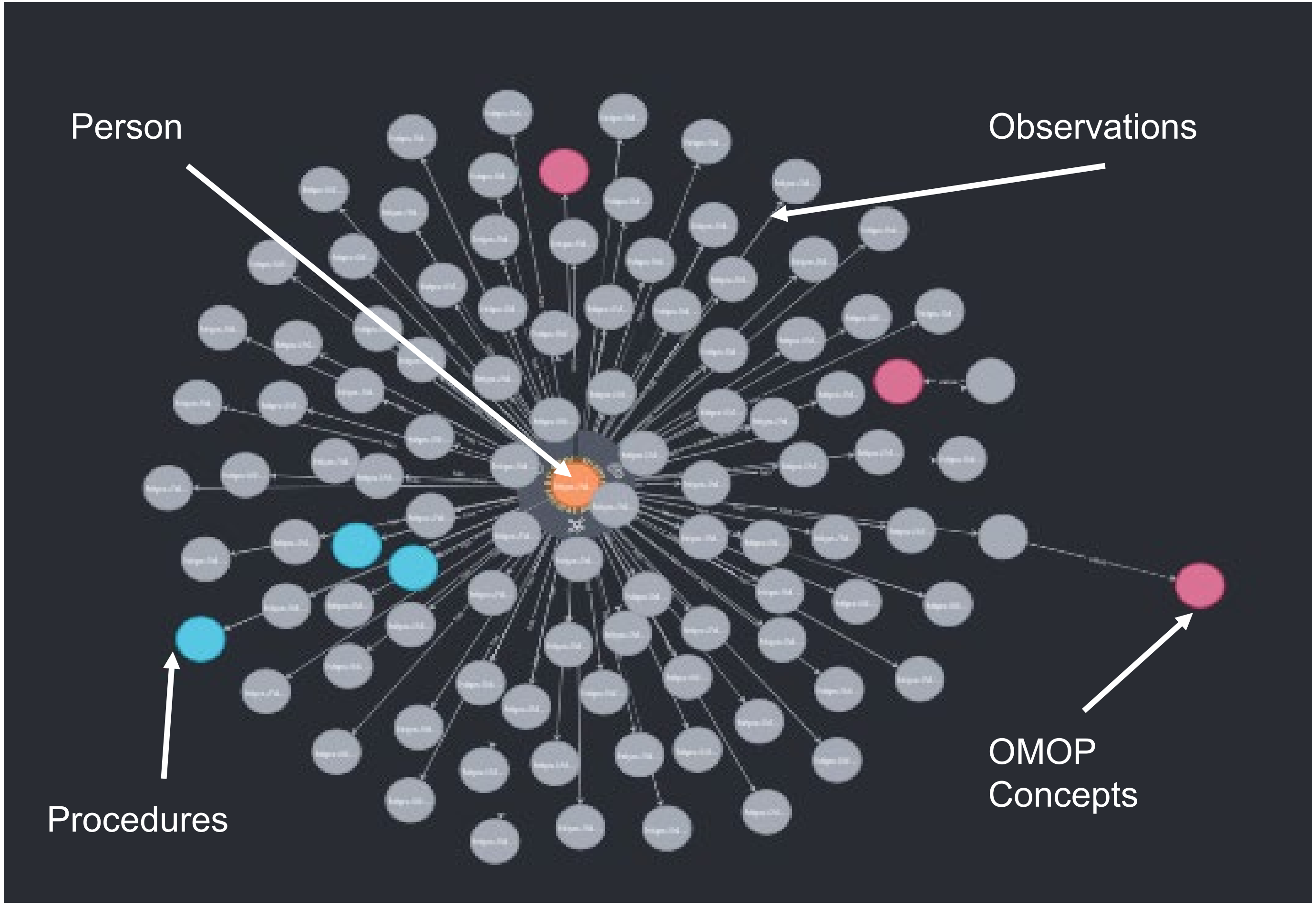
Data Suppliers
Providers
Payers
Justice
Education
Social services

Schemas
Shareable, reusable
Decentralized
Maintained by stakeholders

Research Data Commons
Data collection, curation and harmonization



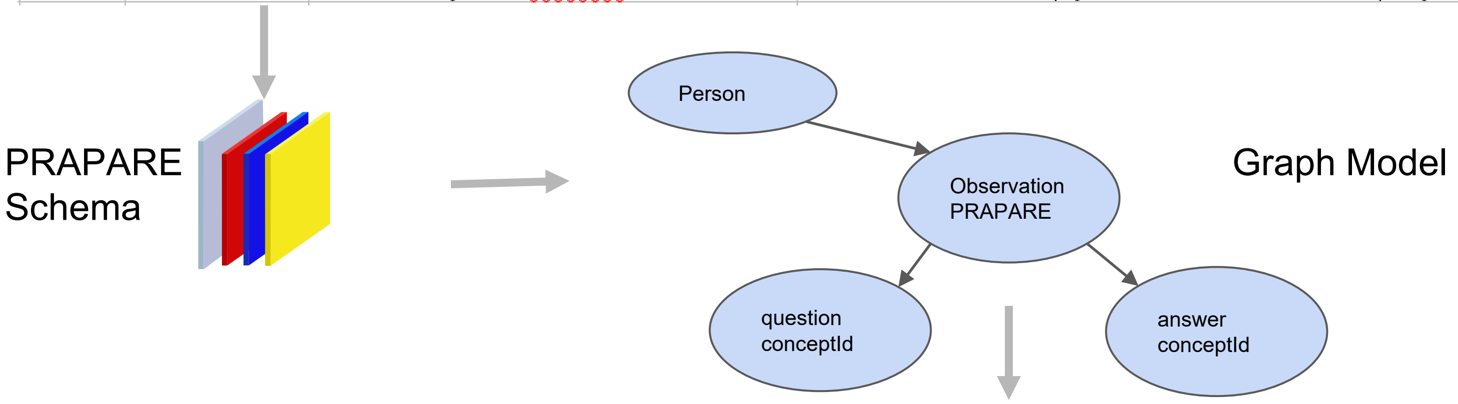
FHIR Patient



Cloud Privacy Labs

OMOP Transformation - PRAPARE

PAT_NBR	QUESTION_DATE	QUESTIONCODE	QUESTIONTEXT	ANSWER
3096135	2020-03-13	Social Integration PRAPARE	Question: "How often do you see or talk to people that you care about and feel close to? Examp	More than 5 times a wee
156546	2020-10-21	Material Security - Food PRAPARE	Questions:"How hard is it to pay for the basics like food", "In the past year have you or any f	N



OMOP Observation

A	B	C	D	E	F	
observation_concept_id	observation_date	observation_datetime	observation_type_concept_id	value_as_string	value_as_concept_id	qual
37020032	2020-03-13	2020-03-13T00:00:00Z		More than 5 times a week	37079490	
37020730	2022-03-30	2022-03-30T00:00:00Z		N	45878245	
46235507	2020-03-13	2020-03-13T00:00:00Z		N	45878245	
37020730	2020-03-13	2020-03-13T00:00:00Z		N	45878245	
37020774	2020-03-13	2020-03-13T00:00:00Z		N		

DARTNet Institute
Informing Practice Improving Care

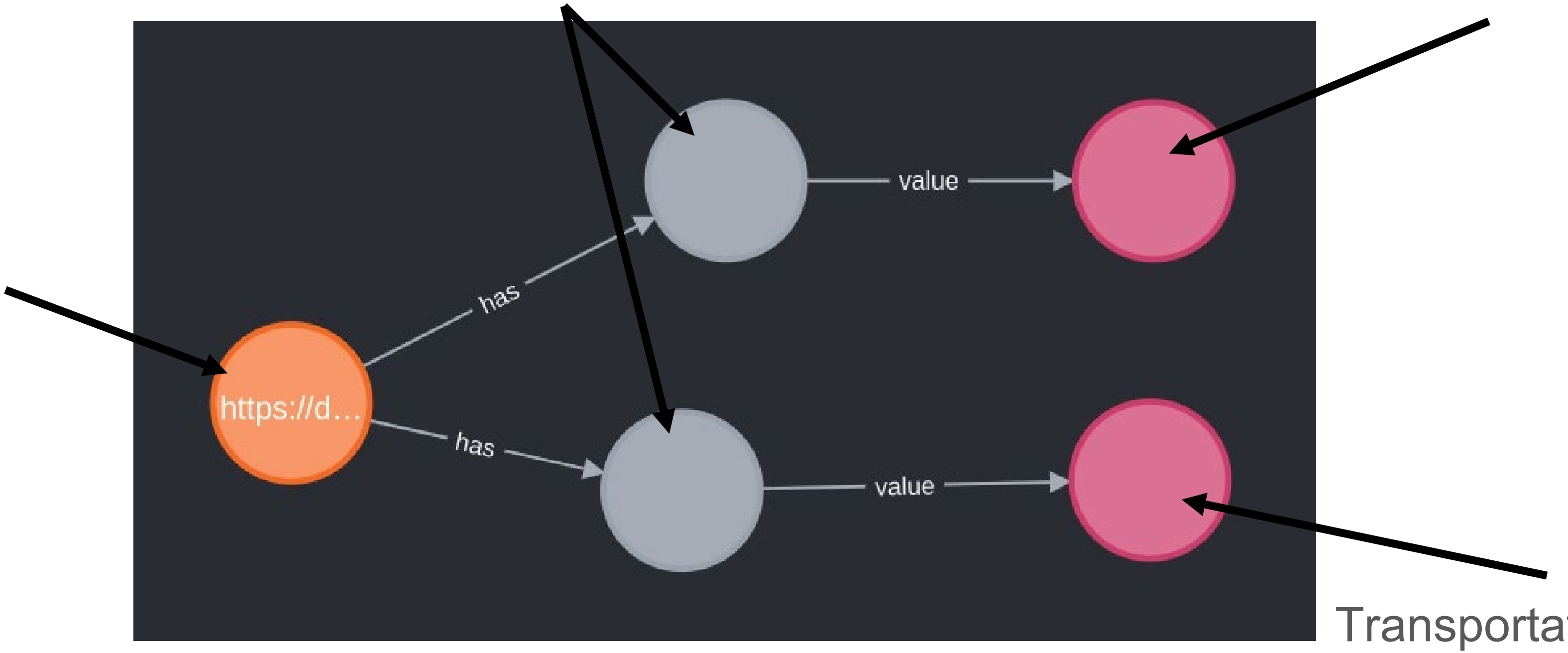
Cloud Privacy Labs

SDoH (PRAPARE)

Observation

Housing
OMOP Concept

Person

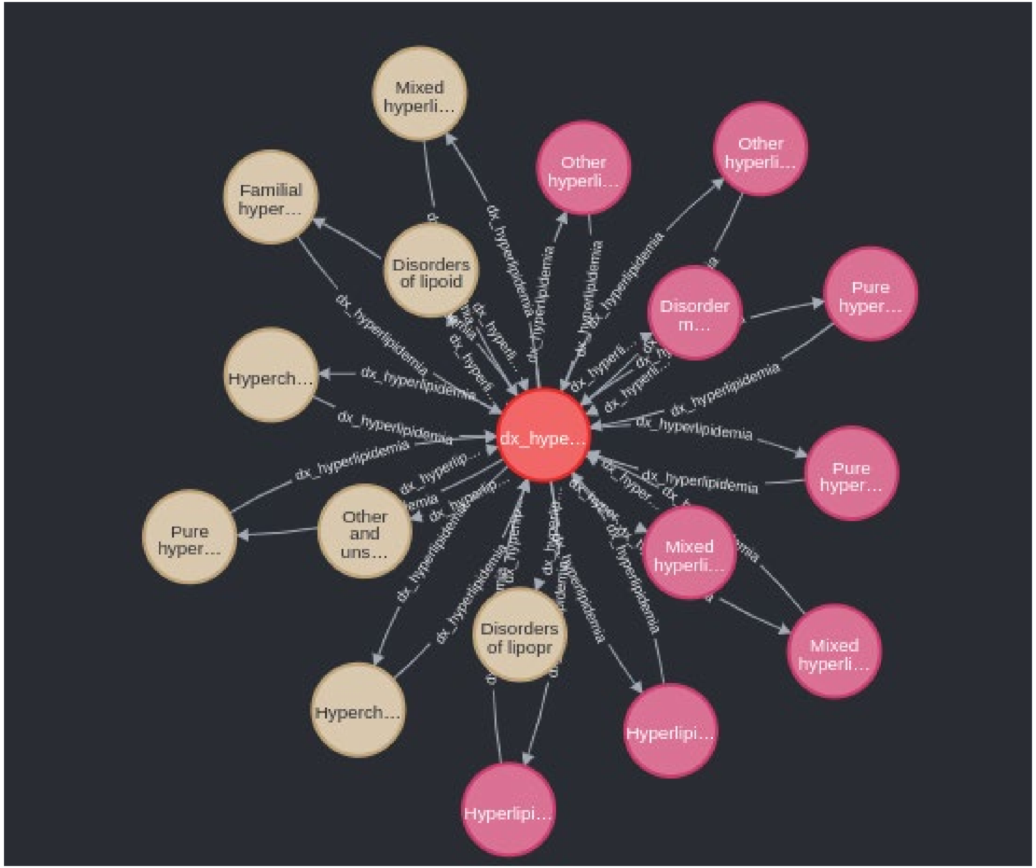


Transportation
OMOP Concept

DARTNet Institute
Informing Practice Improving Care

Cloud Privacy Labs

Valueset (Hyperlipidemia)



A valueset collects related concepts around a single node

Patients cluster around valuesets

Evaluation Overview

- Evaluation 1: Ingestion of data from varying sources using varying formats while maintaining data integrity.
- Evaluation 2: Data reliability, validation and verification of ingestion through the LSA approach compared to standard scripted approaches as represented by DARTNet’s ETL approach to the OMOP CDM.
- Evaluation 3: Evaluation of input for Artificial Intelligence/ Machine Learning compared between the graph database and the OMOP v6 CDM.
- Evaluation 4: Output options: Output to OMOP v5.4 using both ETL processes, evaluate final outputs for similarity.

DARTNet Institute
Informing Practice Improving Care

Cloud Privacy Labs

Evaluations Overview:

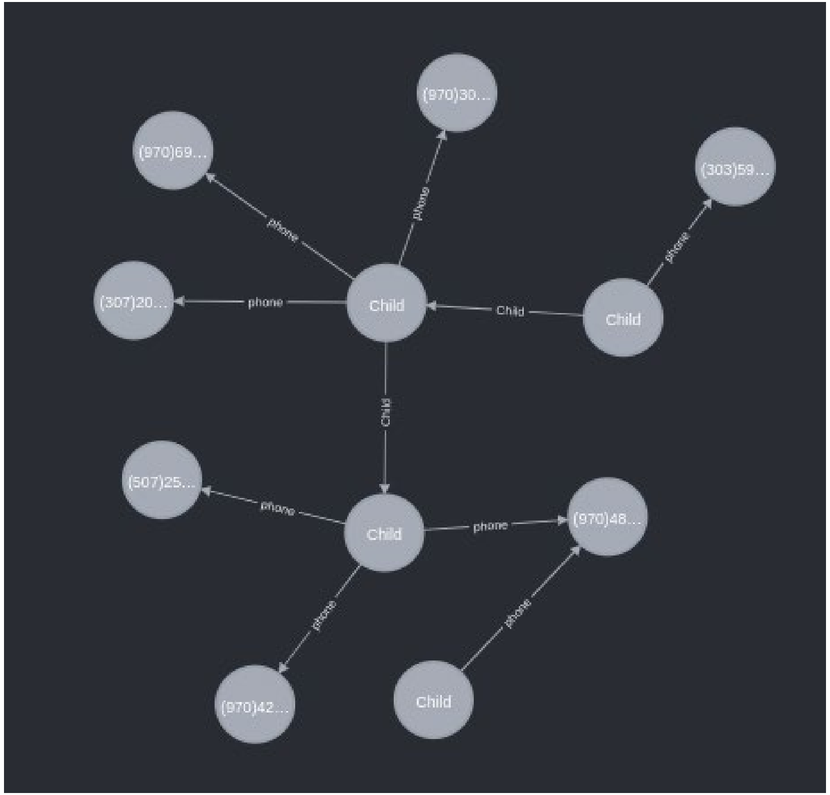
- LSA allows one ingestion with multiple outputs – research CDMs or varying organizational approaches to data ingestion and interpretation
- LSA allows data manipulation and creation of new “data constructs” that can be stored for re-use and easily ingested into AI/ML models
- LSA easily maintains provenance which may relate to reliability, biases, security and complex data linkages
- LSA data ingestion is slower than standard sql scripting but does more at the initial step with multiple output possibilities

DARTNet Institute
Informing Practice Improving Care

Cloud Privacy Labs

Family Linkages – LSA Neo4j Results

- Nodes created for phones and guarantors, relationships displayed in the linkage information
- Future work would add a “family node” to link around
- Information self-aggregated
- In full use more complex linkage data would be used



Evaluation 3: AI Process

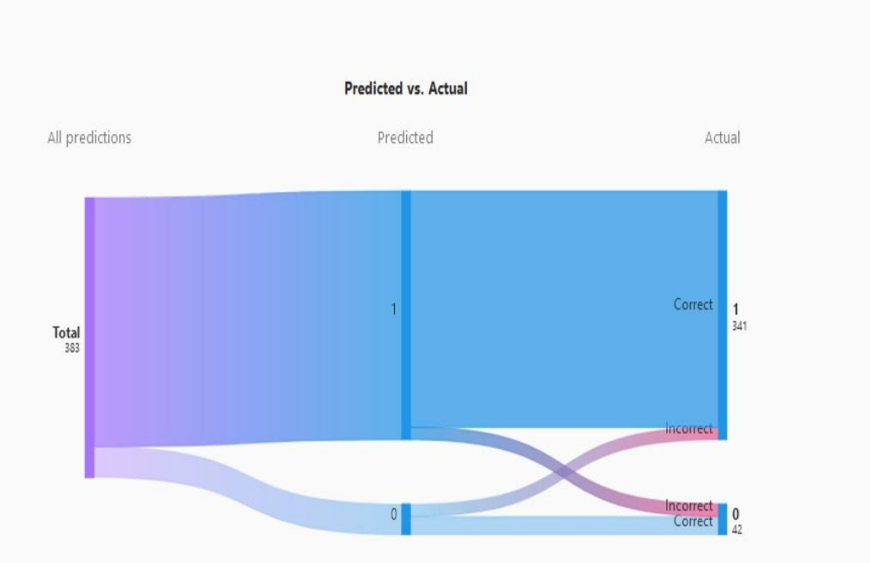
Process:

- OMOP Data frame – 38 variables
- **New variables and calculated variables added to data frame – not to DB**
- Neo4j Data frame – 40 variables
- **Able to easily add new variables or calculated variables into data model**
- Binary outcome model ran for each
- Both models excellent ROC/AUC metrics

DARTNet Institute
Informing Practice Improving Care

Cloud Privacy Labs

Evaluation 3: Model Scoring



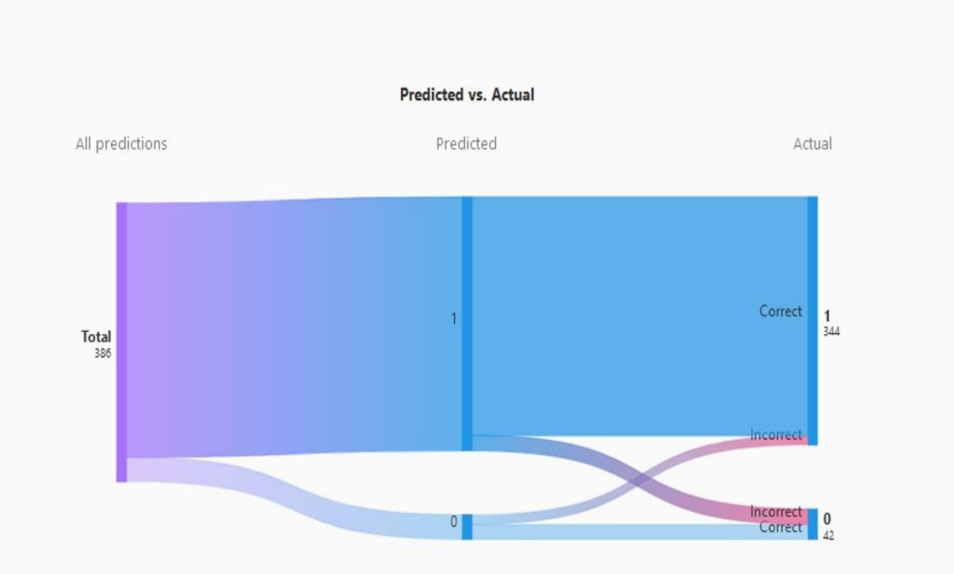
Model accuracy insights

Advanced metrics

If the model predicts **1**, it is correct **95.015% of the time.**

For the values that are **1** in the dataset, the model **predicted** **95.015%** of them to be **1.**

Neo4j



Model accuracy insights

Advanced metrics

If the model predicts **1**, it is correct **94.034% of the time.**

For the values that are **1** in the dataset, the model **predicted** **96.221%** of them to be **1.**

OMOP

DARTNet Institute
Informing Practice Improving Care

Cloud Privacy Labs

Next Steps

- Improve direct input to AI and NLP
- Add multiple output schemas
- Add AI aided schema generation
- Explore NLP as a codification/meta data generator during ingestion
- Ongoing optimization of ingestion across schemas
- Improve ease of use for dissemination
- Standardization of data modeling for health data
- Expansion of the metadata dictionary

Questions?

Bserdar@cloudprivacylabs.com
Wilson.Pace@DARTNet.info

ONC 2023

ANNUAL MEETING

Questions?