Characterizing Respiratory Symptoms and COVID-19 Trends from OMOP-CDM database for Public Health Reporting

Marcello Chang, MS[1*], Jonathan Lu, MS[1*], Birju Patel, MD, MPH[1], Nigam H. Shah, MBBS, PhD[1], Jonathan H. Chen, MD, PhD[1]
[1]Center for Biomedical Informatics Research, Stanford School of Medicine, Stanford, CA

Introduction

- CDC mandates COVID-19 case reporting but:
  - Lack info on patient demographics
  - Interoperable health information infrastructure needed
- We used our clinical data warehouse and common data model to answer the CDC's request

"To understand the proportion of patients who present with acute respiratory illnesses being clinically tested for COVID-19 over time across age, sex, ethnicity and race"

Methods

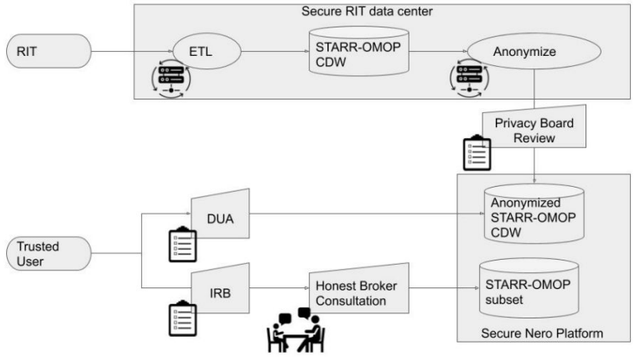


Figure S1: Data governance for the STARR-OMOP dataset (borrowed with permission from 1)

run_date	genderConcept	variable	value	Range
2020-07-21	FEMALE	ARIVisitInTimeRange	24135	['2020-01-01', '2020-07-21']
2020-07-21	MALE	ARIVisitInTimeRange	20090	['2020-01-01', '2020-07-21']
2020-07-21	No matching concept	ARIVisitInTimeRange	15	['2020-01-01', '2020-07-21']
2020-07-21	FEMALE	SARSCoV2TestedInTimeRange	46859	['2020-01-01', '2020-07-21']
2020-07-21	MALE	SARSCoV2TestedInTimeRange	32290	['2020-01-01', '2020-07-21']
2020-07-21	No matching concept	SARSCoV2TestedInTimeRange	51	['2020-01-01', '2020-07-21']
2020-07-21	FEMALE	SARSCoV2DetectedInTimeRange	1358	['2020-01-01', '2020-07-21']
2020-07-21	MALE	SARSCoV2DetectedInTimeRange	1412	['2020-01-01', '2020-07-21']
2020-07-21	No matching concept	SARSCoV2DetectedInTimeRange	5	['2020-01-01', '2020-07-21']
2020-07-21	FEMALE	firstorLastARItoFirstSARSCoV2TestedInDayRange	5860	[-1, 14]
2020-07-21	MALE	firstorLastARItoFirstSARSCoV2TestedInDayRange	4541	[-1, 14]
2020-07-21	No matching concept	firstorLastARItoFirstSARSCoV2TestedInDayRange	11	[-1, 14]
2020-07-21	FEMALE	firstorLastARItoFirstSARSCoV2DetectedInDayRange	598	[-1, 14]
2020-07-21	MALE	firstorLastARItoFirstSARSCoV2DetectedInDayRange	612	[-1, 14]
2020-07-21	No matching concept	firstorLastARItoFirstSARSCoV2DetectedInDayRange	0	[-1, 14]

Example Data Delivery

- Stanford Medicine STARR-OMOP[1] de-identified data warehouse, in OMOP-CDM schema[2], accessed via Google BigQuery.
  - Rapid Prototyping
  - Interoperable (Common Data Model)
  - Data Transfer Agreement
- Visits related to respiratory illnesses (ICD10CM codes)
- SARS-CoV2 NAAT tests and results
- Age, gender, race, ethnicity
- Metadata

Results

- 5 data updates: 6/10/2020 - 7/21/2020
  - 44,240 patients with respiratory symptoms
  - 79,200 patients receiving SARS-CoV2 tests
  - Update in 3 min, re-run on-demand
- tinyurl.com/stanfordcovidcdcmmwr

Table Design

	Age <5 yrs	Age 5-17 yrs	Age 18-49 yrs	Age 50-64 yrs	Age 65-74 yrs	Age 75-84 yrs	Age 85+ yrs
Total visits	6445	8468	13934	6537	3408	1816	453
No. tested	1495	2944	31214	13453	8190	4428	995
No. positive	28	82	1055	470	206	127	50

Wide Table

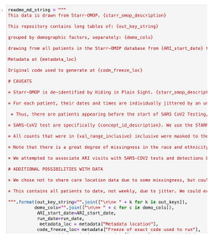
age	variable	value	Range
[00, 05)	ARIVisitInTimeRange	6445	['2020-01-01', '2020-06-28']
[05, 18)	ARIVisitInTimeRange	8468	['2020-01-01', '2020-06-28']
[18, 50)	ARIVisitInTimeRange	13934	['2020-01-01', '2020-06-28']
[00, 05)	SARSCoV2TestedInTimeRange	1495	['2020-01-01', '2020-06-28']
[05, 18)	SARSCoV2TestedInTimeRange	2944	['2020-01-01', '2020-06-28']
[18, 50)	SARSCoV2TestedInTimeRange	31214	['2020-01-01', '2020-06-28']
[00, 05)	SARSCoV2DetectedInTimeRange	28	['2020-01-01', '2020-06-28']
[05, 18)	SARSCoV2DetectedInTimeRange	82	['2020-01-01', '2020-06-28']
[18, 50)	SARSCoV2DetectedInTimeRange	1055	['2020-01-01', '2020-06-28']

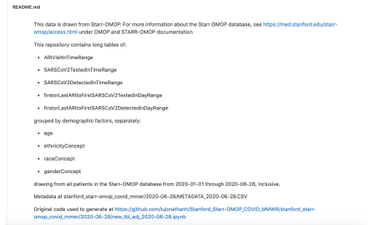
Long Table

ARI & COVID associations

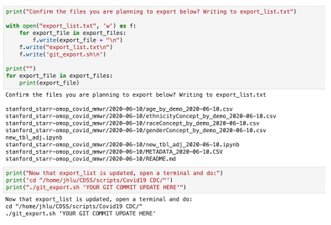
age	variable	value	range
[00, 05)	firstorLastARItoFirstSARSCoV2TestedInDayRange	199	[-1, 14]
[05, 18)	firstorLastARItoFirstSARSCoV2TestedInDayRange	369	[-1, 14]
[18, 50)	firstorLastARItoFirstSARSCoV2TestedInDayRange	4091	[-1, 14]
[00, 05)	firstorLastARItoFirstSARSCoV2DetectedInDayRange	18	[-1, 14]
[05, 18)	firstorLastARItoFirstSARSCoV2DetectedInDayRange	42	[-1, 14]
[18, 50)	firstorLastARItoFirstSARSCoV2DetectedInDayRange	451	[-1, 14]

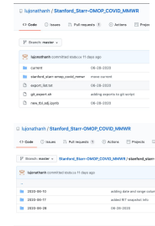
Single Point of Truth





Automated with Traceback





Conclusion

- OMOP-CDM served urgent public health need by facilitating rapid, iterative reporting for clinical trends during a pandemic.
- Limitations: Missing values (workflow/transformation issues?)

References

1. Datta S, Posada J, Olson G, Li W, O'Reilly C, Balraj, D, Mesterhazy J, Pallas J, Desai P, Shah NH. A new paradigm for accelerating clinical data science at Stanford Medicine. Arxiv preprint.
2. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong ICK, Rijnbeek PR, Lei J van der, Pratt N, Norén GK, Li Y-C, Stang PE, Madigan D, and Ryan PB. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform. 2015; 216: 574–578