MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

@horcle_buzz

A symptom-based natural language processing surveillance pipeline for post-COVID-19 patients

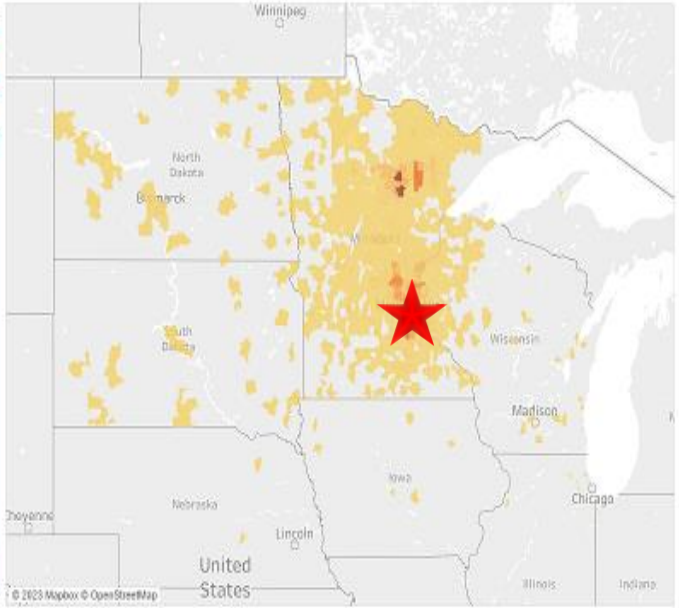
Greg M. Silverman

Senior Systems Developer

University of Minnesota Department of Surgery

Post Acute Sequalae of Sars Cov-2 (PASC)

- Patients with PASC, also known as "long COVID," suffer chronic depression and anxiety, persistent cough, extreme fatigue and other debilitating symptoms that can persist for months [Soriano, *et al.*; Parker, *et al.*; Abdelwahab, *et al.*].
- Globally, up to 65 million people may have PASC [CIDRAP].
- Diagnosis of PASC is difficult:
  - It is not well understood
  - Relevant data in EHR are hidden in free text of clinical notes



OBJECTIVE

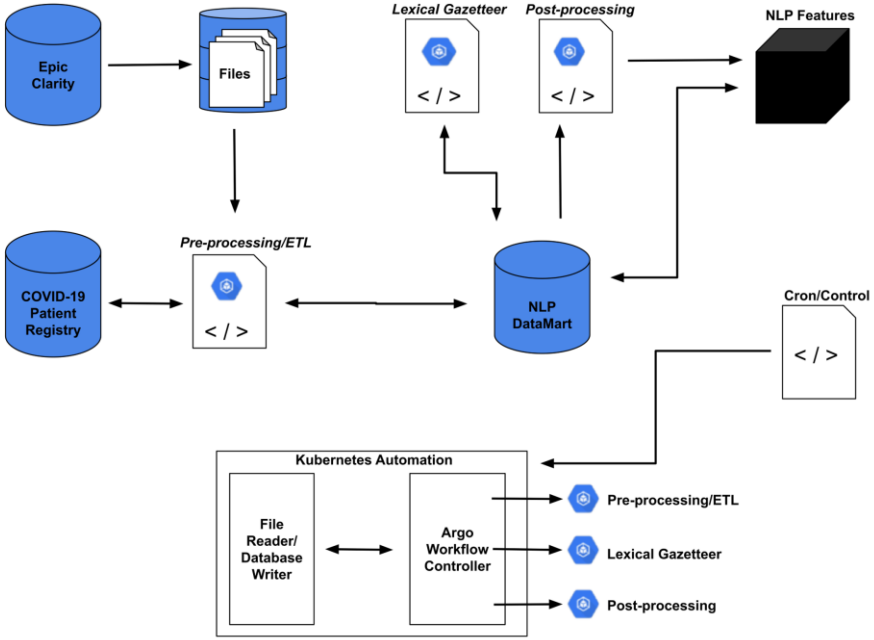
Leverage Natural Language Processing (NLP) to help identify patients at risk for developing PASC to get them referred to a post-COVID-19 clinic for screening.

This study was conducted in the MHealth Fairview network, which includes 12 U.S. hospitals and 60 primary care clinics affiliated with the University of Minnesota.

Cohort

- Inclusion criteria:
  - March 2020 – November 2022
  - Did not opt out of research
  - PCR positive for COVID-19 in MHealth Fairview system (with ICD10 code: U07.1)
- Clinical notes processed:
  - Outpatient encounter notes: 2,237,275
  - Emergency Department Provider notes: 186,037

NLP



NLP Automation Pipeline

Symptom mentions in clinical notes are captured using the rule-based system introduced and validated by Sahoo, *et al.*

- Lexicon was developed using methods outlined and validated in Silverman, *et al.*; Sahoo, *et al.*

PASC Classification: Methods

- 153 randomly selected cases from our cohort that were at least 18 years of age and had symptom mentions in their encounter notes consistent of PASC were reviewed by 3 clinicians.
- Cases were classified for PASC as: "possible"; "unlikely"; "indeterminate."
- Symptoms extracted from notes provided for classification of each case were categorized using the following timeframes:
  - Baseline: All encounters from January 2019 to 14 days prior to infection
  - Acute: All encounters +/- 14 days within date of positive PCR test
  - Post Acute: All encounters (30-60; 60-180; 180-360; 360+) days post-infection

Analysis of Associations: Methods

- Odds ratios (OR) were calculated to determine the risk of having a PASC diagnosis (i.e., ICD10 code of U09.9). Independent variables of interest were:
  - having symptom mentions consistent with PASC (suspected PASC)
  - sex
  - Race/ethnicity
  - age
  - Elixhauser comorbidity index
- We hypothesize that there is a strong association between having PASC symptom mentions and being diagnosed with PASC.

Demographics

			PASC Status (suspected*)	
		Overall	Negative	Positive
n		93446	63115	30331
Age, n (%)	< 40 years	42046 (45.0)	29648 (47.0)	12398 (40.9)
	>= 40 and < 55 years	19255 (20.6)	12690 (20.1)	6565 (21.6)
	>= 55 and < 65 years	13511 (14.5)	8685 (13.8)	4826 (15.9)
	>= 65 and < 80 years	12696 (13.6)	8035 (12.7)	4661 (15.4)
	>= 80 years	5938 (6.4)	4057 (6.4)	1881 (6.2)
Sex, n (%)	female	52943 (56.7)	34667 (54.9)	18276 (60.3)
	male	40490 (43.3)	28439 (45.1)	12051 (39.7)
Race, n (%)	Asian	4395 (5.1)	2894 (5.1)	1501 (5.1)
	Black	11113 (12.9)	7380 (12.9)	3733 (12.8)
	Declined	4950 (5.7)	3910 (6.9)	1040 (3.6)
	Hispanic	2960 (3.4)	1937 (3.4)	1023 (3.5)
	Other	1184 (1.4)	802 (1.4)	382 (1.3)
	White	61610 (71.5)	40096 (70.3)	21514 (73.7)

* Having symptom mentions consistent with PASC

Figure 1a - Distribution of Patients with Symptoms Consistent with PASC

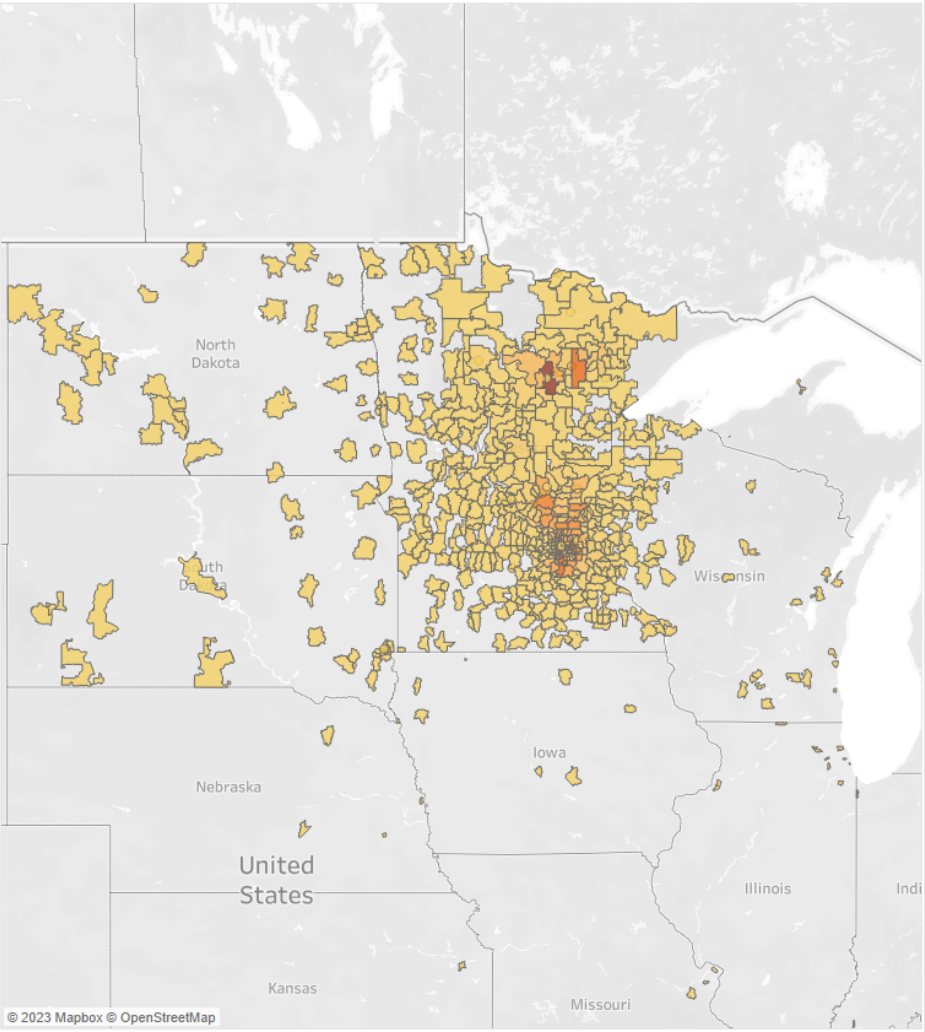
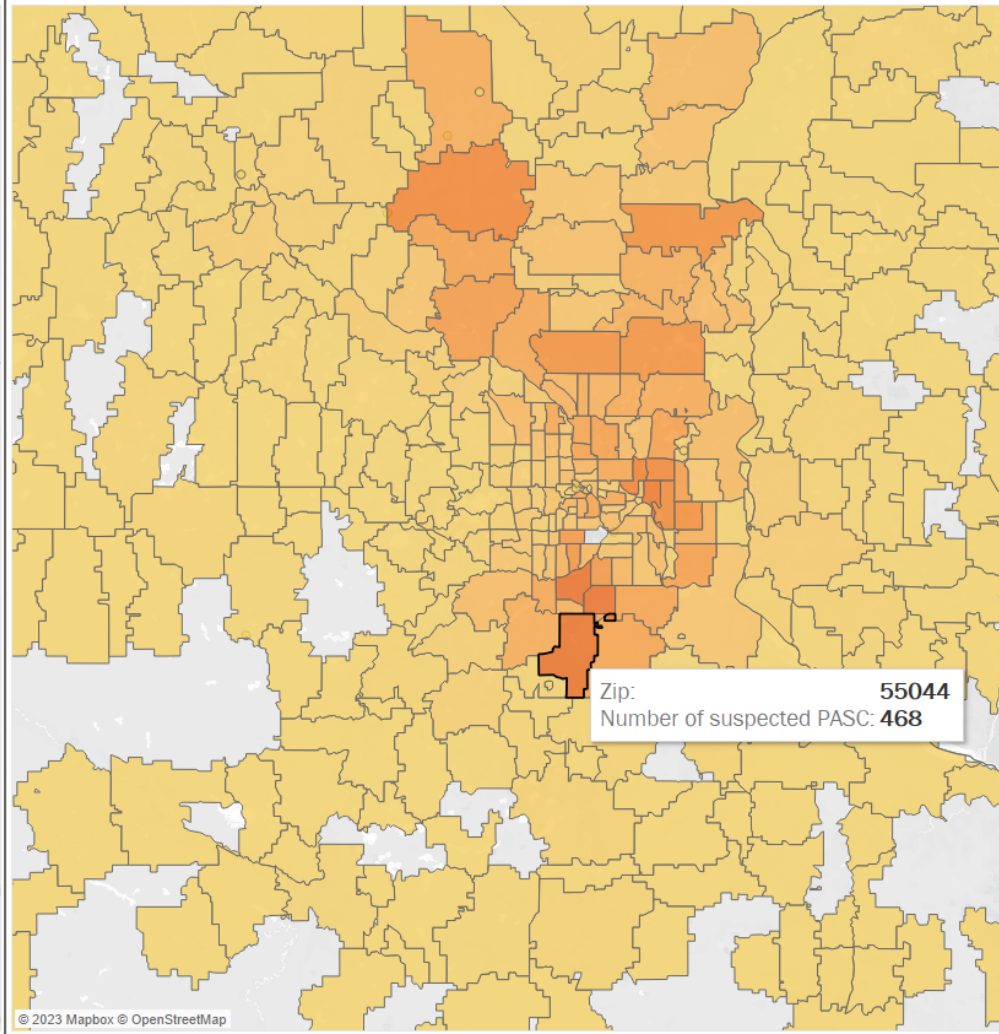


Figure 1b - High Concentration of Patients with Symptoms Consistent with PASC



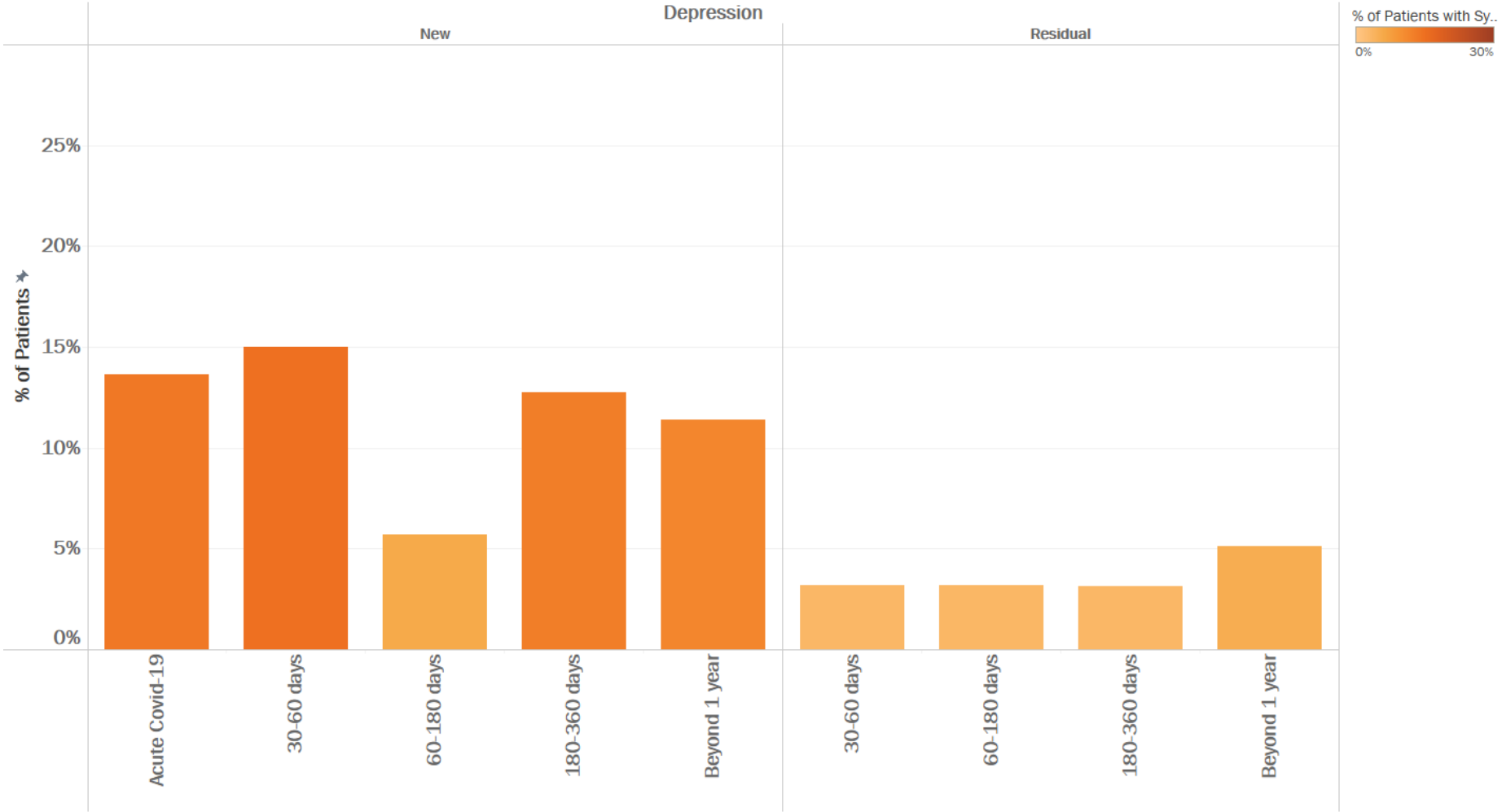
Top Post-Acute COVID-19 Symptom Mentions

New (post-COVID)*	Residual (post-COVID)**
Headache (13123)	Depression (4224)
Depression (12924)	Anxiety (4191)
Nausea/Vomiting (12547)	Headache (3635)
Cough (11839)	Dyspnea (3236)
Fatigue (11640)	Nausea/Vomiting (3230)
Anxiety (11314)	Cough (2905)
Dyspnea (10990)	Fatigue (2737)
Palpitation (10217)	Fever (2182)
Fever (9761)	Palpitation (2176)
Skin rash (8209)	Skin rash (1970)

* New symptoms occurring at least 30 days AFTER acute infection

** Residual symptoms from acute infection persisting over time

New and Residual Mentions of Depression for Males over Time



Mentions of Symptoms by Sex over time: Group: All. Selected symptom: Depression.

PASC Classification: Results

Case summary:

- 44% unlikely PASC
- 43% indeterminate
- 13% possible PASC

PASC Diagnosed (as of June 13, 2023)

COVID-19 Dx Year	Post-COVID-19 Suspected PASC**	Post-COVID-19 Suspected PASC** (with ICD10 code of U09.9)
2020	18894	259
2021	9165	215
2022	3211	72
2023	89	3

* Symptom mentions consistent of PASC

Analysis of Associations: Results

- Being diagnosed with PASC (having an ICD10 code of U09.9) had:
  - An increased risk given:
    - One or more suspected PASC symptoms (OR 3.40, p-value < 0.001)
    - One on more comorbidities (OR 1.15, p-value < 0.001)
  - A reduced risk given:
    - Being male (OR 0.84, p-value < 0.05)
    - Being Black (OR 0.75, p-value < 0.05)
  - Other independent variables that were not significant

Discussion

- While NLP extracted symptoms helped to rapidly assess patients for risk of PASC, gaps in encounter data led to a high number of cases classified as indeterminate.
  - Thus, prevalence of PASC in Post-COVID-19 cohort is likely underestimated
- There is an increased risk of being diagnosed with PASC when symptom mentions consistent with PASC are present in a patient's clinical notes thereby validating our hypothesis.
  - However, the reduced risk for Black and male patients warrants further examination.

Key Limitations

- Incomplete reporting and limited documentation may hinder accurate assessment.
- Definition of PASC:
  - Diverse symptoms and lack of universal definition make it difficult to establish a definitive diagnosis.
  - Population bias (no at-home testing and multiple infections)

Next Steps

- Extend PASC validation
- Develop more robust symptom progression models by:
  - Expanding definition and lexicon of PASC [Thaweethai, *et al.*; Wang, *et al.*]
  - Using LLMs for general signs and symptoms detection across post-acute illnesses.

In Summary

- NLP can assist clinicians in identifying patients at risk of developing PASC by providing methods to evaluate the progression of patient symptoms, which is imperative for improving outcomes.

Thanks to:

My colleagues

Australasian Institute of Digital Health

Contact information:

- Greg M. Silverman (gms@umn.edu)
- Christopher Tignanelli (ctignane@umn.edu)