A Study of COVID-19 Impact on Mental Health via Information Extraction and Querying of Clinical Notes

COVID CLSAC: ANALYTICS FOR PANDEMIC DECISION SUPPORT

Marie Humbert-Droz, PhD

Center for Biomedical Informatics Research, Stanford University

mhdroz@stanford.edu

Introduction

- **1/20/2020**
  - First COVID-19 case identified in the U.S.

- **1/31/2020**
  - Declaration of public health emergency (U.S.)
  - First confirmed COVID-19 patient in Santa Clara County

- **3/1/2020**
  - Declaration of national emergency

- **3/6/2020**
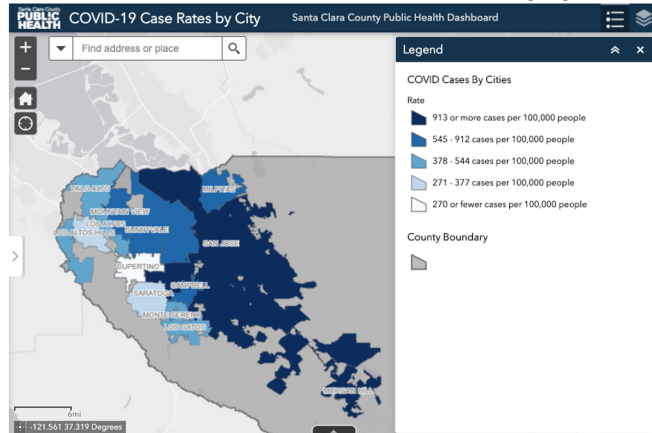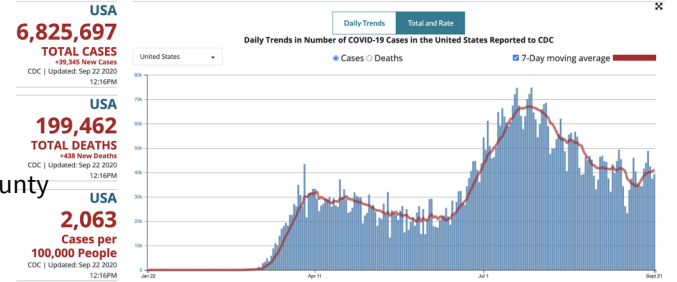 - Stanford Hospital announces it is caring for COVID-19 patients

- **3/9/2020**
  - County announces first death from COVID-19

20

lace order emitted for Santa Clara County

Trends in Number of COVID-19 Cases in the US Reported to CDC, by State/Territory

USA
6,825,697
TOTAL CASES
+39,345 New Cases
CDC | Updated: Sep 22 2020
12:16PM

USA
199,462
TOTAL DEATHS
+438 New Deaths
CDC | Updated: Sep 22 2020
12:16PM

USA
2,063
Cases per
100,000 People
CDC | Updated: Sep 22 2020
12:16PM

*covid.cdc.gov, 7/22/20

- COVID-19 widespread but regions touched differently
- Shelter-in-place order in Santa Clara county was the first in U.S.
- Relatively less admissions at Stanford for COVID-19 compared to UC hospitals
- Syndrome surveillance is primordial for pandemic management
- Address a gap for psychiatric disorders

Stanford University

Introduction – this study

Dataset description	Sept. 1 2019 – June 6 2020
Total number of patients	581,818
Age (years)	
Mean	41
STD	25
Female	323,464 (55.60%)
Male	257,998 (44.34%)
Unknown	356 (0.06%)
Total number of notes	6,550,495
Total number of terms extracted	203,257,129
Total number of encounters	8,431,624
Total number of DX codes	14,783,096

- Extraction of all clinical terms from clinical notes
- Differential analysis of psychiatric disorders mentions
- Comparison with an average over the three years prior
- Analysis of 4 broad age groups

| Age group	Age range (years)	Number of patients
1	< 18	78,486
2	18 - 40	76,535
3	40 - 65	108,097
4	> 65	85,318

Stanford University

Introduction – pipeline and terminology

Raw clinical text → Tokenizer → Phrase Matcher → Context – concept - context → Concept - labels

P* T* E*

* Polarity – Temporality - Experiencer

System Organ Class (SOC)
| Grouping by site, aetiology or purpose | e.g. Skin and subcutaneous tissue disorders |

High Level Group Term (HGLT)
| For data retrieval purposes and to group HLTs together | e.g. Epidermal and dermal conditions |

High Level Term (HLT)
| Descriptive group for subordinate PTs | e.g. Skin injuries and mechanical dermatoses |

Preferred Term (PT)
| Single medical concept | e.g. Contusion |

Lowest Level Term (LLT)
| Colloquial version or synonym for PT, or sub-element | e.g. Bruise |

```
System Organ Class
└── Psychiatric disorders
    ├── Anxiety disorders and symptoms
    │   ├── Anxiety disorders NEC
    │   ├── Anxiety symptoms
    │   │   ├── Anxiety
    │   │   ├── Nervousness
    │   │   ├── Stress
    │   │   └── Tension
    │   ├── Fear symptoms and phobic disorders (incl social phobia)
    │   ├── Obsessive-compulsive disorders and symptoms
    │   ├── Panic attacks and disorders
    │   └── Stress disorders
    │       ├── Acute stress disorder
    │       ├── Burnout syndrome
    │       └── Post-traumatic stress disorder
    ├── Deliria (incl confusion)
    ├── Depressed mood disorders and disturbances
    ├── Manic and bipolar mood disorders and disturbances
    ├── Mood disorders and disturbances NEC
    ├── Psychiatric and behavioural symptoms NEC
    ├── Sleep disorders and disturbances
    └── Suicidal and self-injurious behaviours NEC
```

- ~ 970,000 terms:
  - 27 System Organ classes
  - 337 High Level Group Terms
  - 1,737 High Level Term
  - 23,954 Preferred Term
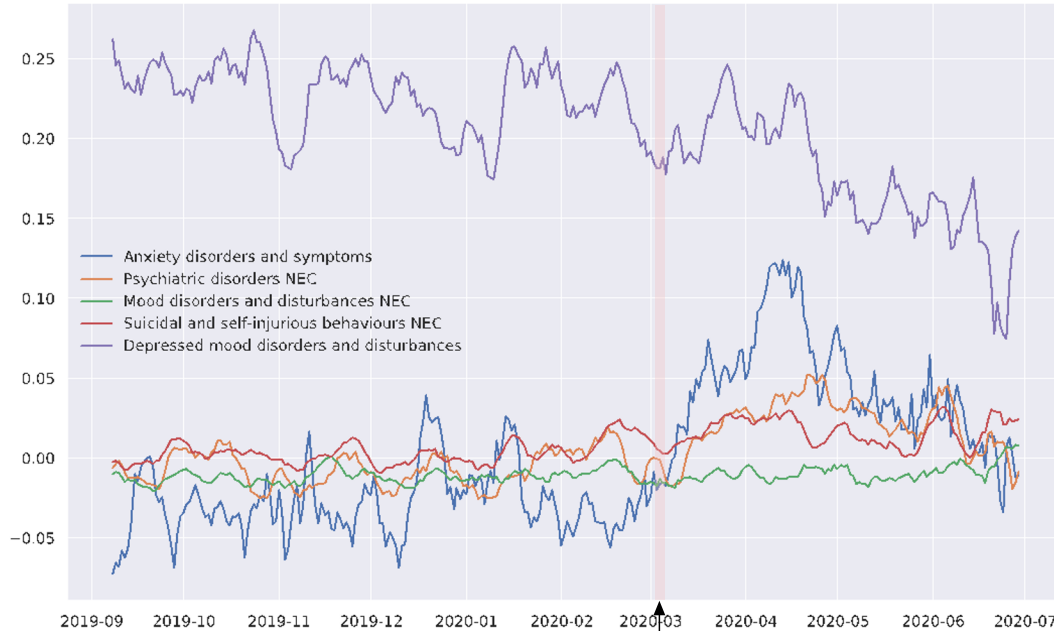 - 71,603 Lowest Level Term

Stanford University

Results – Term counts per unique patient
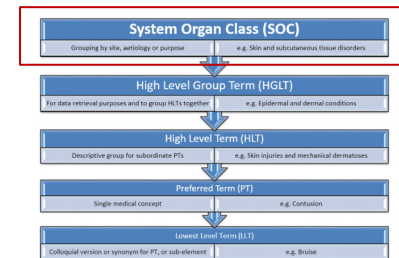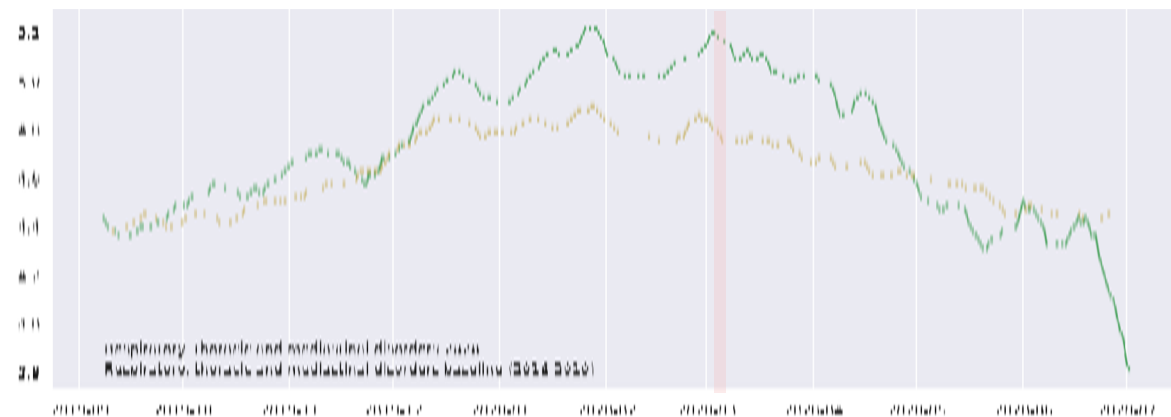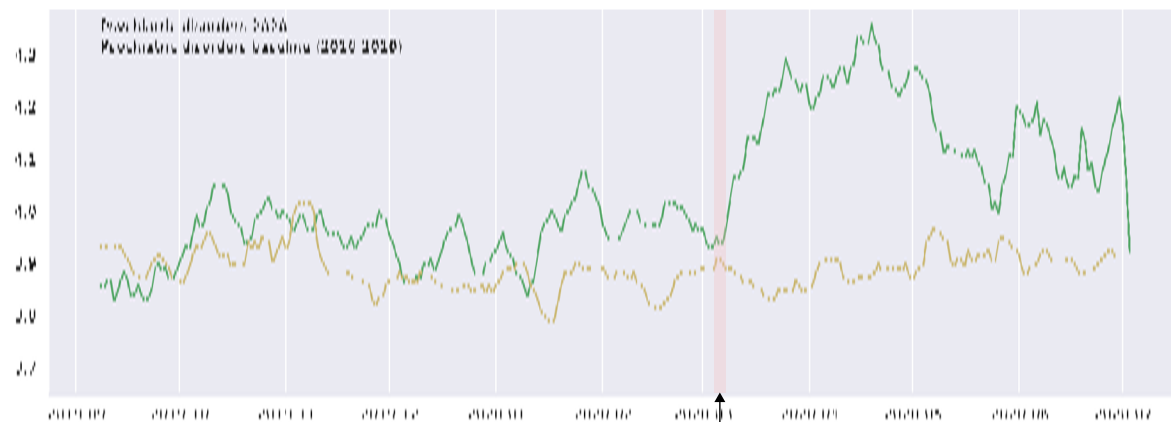
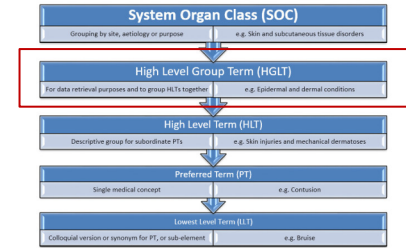March 17 2020: Shelter-in-place order in the San Francisco Bay Area starts

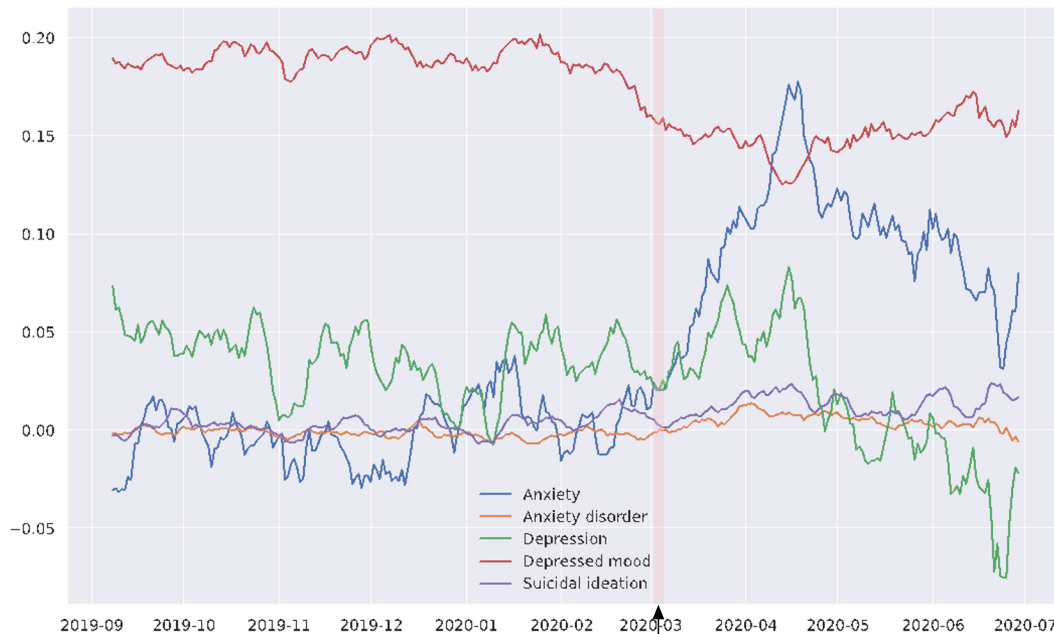* 30 days jitter due to deidentification of the data

Results – Term counts per unique patients: differential with baseline
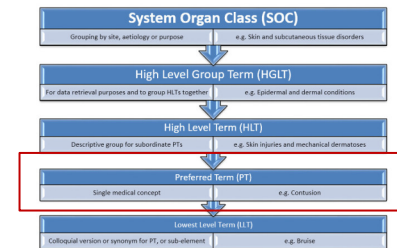

March 17 2020: Shelter-in-place order in the San Francisco Bay Area starts

Results – Term counts per unique patients: differential with baseline

March 17 2020: Shelter-in-place order in the San Francisco Bay Area starts
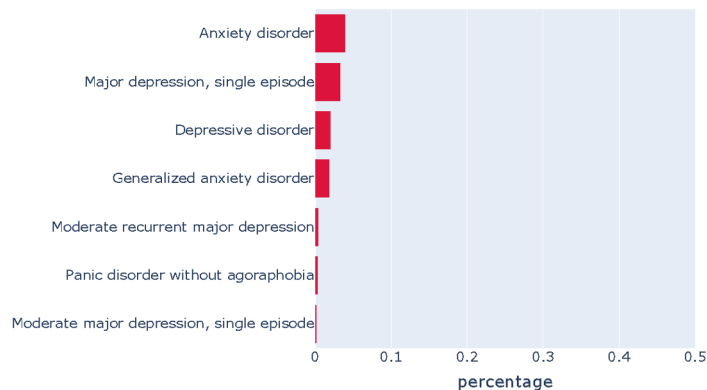
Stanford University

Analysis of psychiatric disorders mentions

- Considering all patients and all notes from Sept 1 2019 to June 6 2020 shows:
  - Sharp increase in psychiatric disorders related terms mentions per patient almost directly after shelter-in-place order is emitted
  - Majority of the signal comes from anxiety symptoms
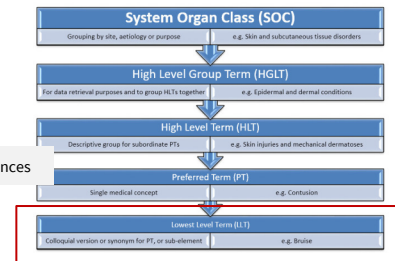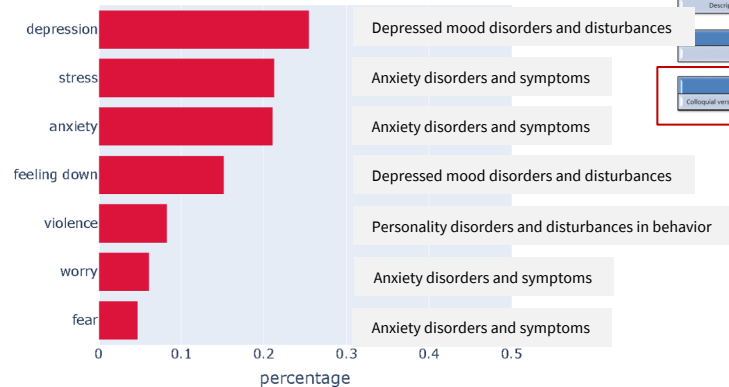
- What about diagnosis codes?

Top mental health DX code vs NLP extractions from notes

Top 7 psychiatric diagnosis codes

- Anxiety disorder
- Major depression, single episode
- Depressive disorder
- Generalized anxiety disorder
- Moderate recurrent major depression
- Panic disorder without agoraphobia
- Moderate major depression, single episode

Top 7 psychiatric mentions

- depression — Depressed mood disorders and disturbances
- stress — Anxiety disorders and symptoms
- anxiety — Anxiety disorders and symptoms
- feeling down — Depressed mood disorders and disturbances
- violence — Personality disorders and disturbances in behavior
- worry — Anxiety disorders and symptoms
- fear — Anxiety disorders and symptoms

System Organ Class (SOC)
Grouping by site, aetiology or purpose | e.g. Skin and subcutaneous tissue disorders

High Level Group Term (HGLT)
For data retrieval purposes and to group HLTs together | e.g. Epidermal and dermal conditions

High Level Term (HLT)
Descriptive group for subordinate PTs | e.g. Skin injuries and mechanical dermatoses

Preferred Term (PT)
Single medical concept | e.g. Contusion

Lowest Level Term (LLT)
Colloquial version or synonym for PT, or sub-element | e.g. Bruise

- Considering diagnosis codes:
  - 4 out of 7 codes relate to depression, 2 to anxiety and 1 to panic
  - 4 out of 7 mentions belong to Anxiety disorders and symptoms group, 2 to Depressed mood disorders and disturbances
  - The proportion of codes is about an order of magnitude lower than the proportion of mentions in the text
  - Wider range of concepts is captured in the text:
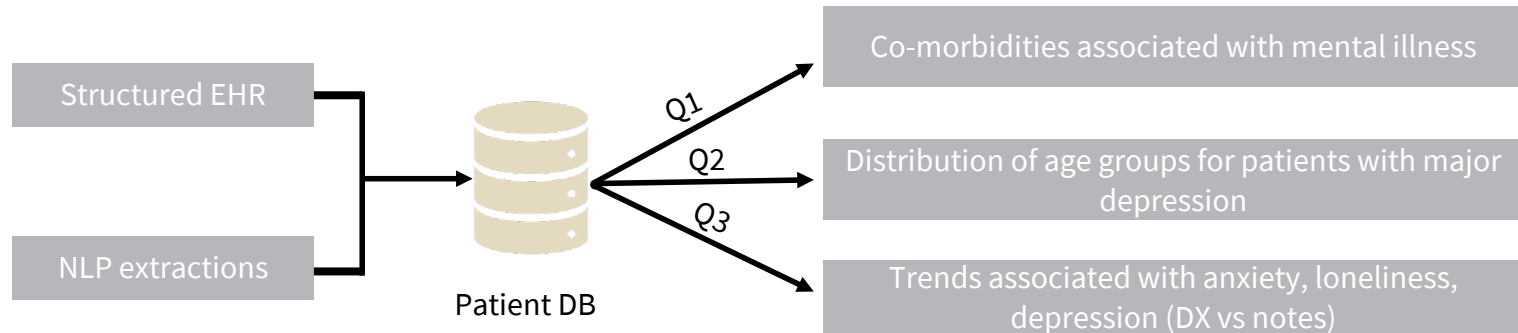   - Feeling down, worry, fear

Monthly trend of top 7 extracted concepts by age group

Top 7 psychiatric terms for all age groups: monthly differentials

Anxiety mentions increase for all age groups after the shelter-in-place order. Fear is felt mostly by the youngest group and worry by the eldest group

Work in progress

- Patient database build for queries
  - Combine information from both structured and unstructured EHR
  - Patient – clinical event data
  - Formulate and answer more complex questions

Structured EHR

NLP extractions

Patient DB

Q1 → Co-morbidities associated with mental illness

Q2 → Distribution of age groups for patients with major depression

Q3 → Trends associated with anxiety, loneliness, depression (DX vs notes)

Conclusions and outlook

- Ontology-based extraction and weakly supervised labeling allows for rapid extraction of all clinical terms from a large hospital EHR database
- Highlight significant trends following shelter-in-place orders:
  - Substantial increase in psychiatric disorders mention
  - Mostly anxiety
  - All considered age groups
- Information extraction from text is primordial to capture such a trend
- Going further:
  - Validation of extraction tool
  - Use of neural language model to improve NER and assertion status
  - Query with embeddings for higher resolution information

Acknowledgements

- Stanford team:
  - Dr. Suzanne Tamang
  - Dr. Pritam Mukherjee
  - Pr. Olivier Gevaert
- PNNL team:
  - Dr. Sutanay Choudhury
  - Colby Ham
  - Kushbu Agarwal