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**
  - Santa Clara County announces first death from COVID-19
  - 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
Introduction – this study

- 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

### Dataset description

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

### Age group distribution

<table>
<thead>
<tr>
<th>Age group</th>
<th>Age range (years)</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt; 18</td>
<td>78,486</td>
</tr>
<tr>
<td>2</td>
<td>18 - 40</td>
<td>76,535</td>
</tr>
<tr>
<td>3</td>
<td>40 - 65</td>
<td>108,097</td>
</tr>
<tr>
<td>4</td>
<td>&gt; 65</td>
<td>85,318</td>
</tr>
</tbody>
</table>
Introduction – pipeline and terminology

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

  * Polarity – Temporality – Experiencer

  ~ 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

---

**System Organ Class (SOC)**
- Grouping by site, etiology or purpose: e.g. Skin and subcutaneous tissue disorders
- High Level Group Term (HGLT): For data retrieval purposes and to group HTTs 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. Confusion
- 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
  - 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
    - Post-traumatic stress disorder
    - Delirium (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
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
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

- 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

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 → Patient DB:
- Q1: Distribution of age groups for patients with major depression
- Q2: Co-morbidities associated with mental illness
- Q3: Trends associated with anxiety, loneliness, depression (DX vs notes)

NLP extractions
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