VA COVID-19 Shared Data Resource
Outline

- Review the VA COVID-19 Shared Data Resource (CSDR)
- Introduction to Natural Language Processing (NLP)
- Current and upcoming CSDR phenotypes supplemented by NLP

**Phenotype**: any clinical condition or characteristic that can be queried from the EHR
Collaboration
Pre-Index Observations

30 days

2 years

Index Date

Post-Index Observations

7 days

60 days
CSDR documentation and support

- Internal wiki pages
- Data access instructions
- Customer support
- Power BI reports

COVID-19:Shared Data Resource - VA Phenomics Library
NLP project workflow

1. Research idea
2. Research design
3. Collecting data
4. Analyzing data
5. Publication process

Input/output data specifications
Concept definition

NLP project
- Planning
- Analysis
- Design
- Implementation
- Testing
- Deployment

NLP data output

Image Courtesy: Olga Patterson
Phenotypes supplemented through NLP

• Vital signs
• Reason for stay
• Symptoms
• Mechanical ventilation
• Supplemental oxygen
• Menopause
• Negative COVID-19 tests
Vital Signs


VITAL SIGNS: Temperature 98.1, pulse 64, respirations 20, blood pressure 144/69, height 6’1”, weight 229 pounds, saturation O2 stats are 97 percent on room air.

VITALS:
T: 98.2
RR: 20
Pulse: 88
BP: 126/50
Weight: 217.2
Height: 66
BMI: 35.1

Physical Exam
Vitals 97.9 F [36.6 C]
70 in [177.8 cm]
229 lb [104.1 kg]

VS 152/94 98.3 85 16

VS 150-160/90-100 98-98.5 80-85 16-22

F1 score: 0.880
## Reason for Stay

- Inpatient.[AdmitDiagnosis]
- EDISLog.[PatientVisitReason]
- EDISDischargeDiagnosis.[DischargeDiagnosis]

### COVID-19: ORDCovid PreIndexSymptoms - VA Phenomics Library

<table>
<thead>
<tr>
<th>Admit Diagnosis</th>
<th>Snippets</th>
<th>CUI</th>
<th>STR</th>
<th>STY</th>
<th>polarity</th>
<th>uncertainty</th>
<th>HistoryOf</th>
</tr>
</thead>
<tbody>
<tr>
<td>n/v/d nausea</td>
<td>nausea</td>
<td>C0027497</td>
<td>Nausea</td>
<td>Sign or Symptom</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>n/v/d vomiting</td>
<td>vomiting</td>
<td>C0042963</td>
<td>Vomiting</td>
<td>Sign or Symptom</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>n/v/d diarrhea</td>
<td>diarrhea</td>
<td>C0011951</td>
<td>Diarrhea</td>
<td>Sign or Symptom</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient Visit Reason</th>
<th>Snippets</th>
<th>CUI</th>
<th>STR</th>
<th>STY</th>
<th>polarity</th>
<th>uncertainty</th>
<th>HistoryOf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si, Detox, Fevers, SOB</td>
<td>detoxication</td>
<td>C0150543</td>
<td>Detoxification procedure</td>
<td>Therapeutic or Preventive Procedure</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Si, Detox, Fevers, SOB</td>
<td>suicidal ideation</td>
<td>C0424000</td>
<td>Feeling suicidal finding</td>
<td>Finding</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Si, Detox, Fevers, SOB</td>
<td>fever</td>
<td>C0019567</td>
<td>Fever</td>
<td>Sign or Symptom</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Si, Detox, Fevers, SOB</td>
<td>shortness of breath</td>
<td>C0013404</td>
<td>Dyspnea</td>
<td>Sign or Symptom</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

F1 score: 0.872
Coronavirus Disease 2019 (COVID-19) Screen: <<COVID_SCREEN>>

The patient reports that they do not have NEGATED_EXISTENCE a fever FEVER.

The patient reports that they do not have NEGATED_EXISTENCE a new or worsening cough COUGH or shortness of breath SOB.

The patient reports they do not have NEGATED_EXISTENCE any flu-like symptoms COLDFLU.

Results: <<RESULTS>> Screen is negative.

COVID-19:ORDCovid PreIndexSymptoms - VA Phenomics Library

F1 score: 0.964
Identifying the need for NLP – Mechanical Ventilation

Structured procedure data for intubation Tier 2
- ICD10 Proc CPT4
  - Mechanical vs. non mechanical
  - Potentially missing instances?
- Intubation codes present when ventilation codes are not present
- Potential delay in data entry?

Structured data for intubation medications Tier 3
- LocalDrugName
  - Ever/never ventilation is fairly well captured with procedure and medication data
  - Start date is the first event, but what about end date (i.e., date of extubation)?

Natural Language Processing Tier 1
- Indication of extubation consistently exits in clinical notes
- End date (End-Start= duration of mechanical ventilation)
- What about patients who were intubated more than once during the same stay?

Natural Language Processing Tier 2
- Multiple intubation events are isolated
- Reintubations and intubations from a new admission
- Are all instances of mechanical ventilational equal?

Natural Language Processing Tier 3+
- TIU
- TIU and conditions and procedures

Classification/disambiguation of emergent vs. procedural intubations
- Other details about ventilation episodse (e.g., prone)

Image Courtesy: Kristine Lynch, VINCI
Mechanical Ventilation

VA EHR Ventilation Phenotype Logic

All COVID patients input

Structured Events

Rule-Based NLP Events

Ventilation Events

Sequence Logic

Procedure Classification

Validation at Each Stage of Development

NLP Intubations: 96% Precision
NLP Extubations: 97% Precision

Identification of Ventilated Patients (Ever/Never)*:
Accuracy: 94.4%
PPV: 97.28
Sensitivity: 93.46%**

Classification of Elective Procedure Intubations:
96% Precision
90% Recall

<table>
<thead>
<tr>
<th>PatientICN</th>
<th>VentilatorStartDate</th>
<th>VentilatorStopDate</th>
<th>VentilationStartType</th>
<th>VentilationStopType</th>
<th>SequenceNumber</th>
<th>ProcedureVentilation</th>
<th>ProneVentilation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2020-05-04</td>
<td>2020-05-08</td>
<td>Intubation Code</td>
<td>NLP Exubation</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2020-05-10</td>
<td>2020-05-14</td>
<td>Medication Code</td>
<td>NLP Exubation</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Image Courtesy: Patrick Alba, VINCI
Identifying the need for NLP - Supplemental Oxygen

• Discrepancy analysis of structured data and NLP
Supplemental Oxygen

- Supplemental oxygen with high flow or low flow nasal cannula, or noninvasive ventilation
- Identified using a combination of structured vitals data and NLP
- Two concepts identified via NLP
  - Oxygen delivery device (“Pt currently on HFNC”)
  - LPM (“Patient’s SpO2 = 98% on 3 L O2”)

- COVID-19: Oxygen Flow (COVID) - VA Phenomics Library
Supplemental Oxygen – PostIndexProcedures

- OxygenLowFlow60d
- OtherUnk60d
- NonInvasiveVentilation60d
- OxygenHighFlow60d
- BloodTransfusion60d
- Dialsy60d
- Contactisol60d
- Dropisol60d
- Airisol60d
- ECMO60d

Count

0K  50K  100K  150K  200K
Hypothetical patient timeline

- Admitted into the hospital on day 1
- Placed in prone position on day 2 with supplemental oxygen
- Intubated and ventilated on day 4
- Successfully extubated on day 7
- Condition worsens on day 10 and they are reintubated
- Successfully extubated again on day 12
- Discharged on day 15
Menopause

- Difficult to classify patients as pre- or postmenopausal on a certain date using structured data
  - Medication orders
  - Coded diagnoses
  - Surgical procedures

- Increased COVID-19 severity in men vs women

- Is there a difference in severity between pre- and postmenopausal women?
Menopause

- The system identifies concept-date pairs
  - Last menstrual period (LMP) date
  - Surgical procedure date
  - Menopause diagnosis date

- Examples
  - “LMP: Feb 15, 2021”
  - “Bilateral oophorectomy 3/2012”
  - “Premature menopause 2015”

- Patient classifications
  - Menopausal, surgical
  - Menopausal, non-surgical
  - Not Menopausal
  - Unknown

- Chart abstraction showed many patients had no documented evidence of menopausal status
Menopause

Patient Classification based on NLP Alone

- Unknown
- Not Menopausal
- Menopausal, surgical
- Menopausal, non-surgical

Number of Patients

AgeAtIndexDate

Image Courtesy: Hannah Eyre, VINCI
Negative COVID-19 Tests

- Veterans are tested for COVID-19 by VA and non-VA providers
- Previously developed NLP system identifies positive COVID-19 tests from non-VA providers, but not negative tests

**Purpose**: identify negative COVID-19 test and test date in clinical notes
Negative COVID-19 Tests

- The minimum for output is a COVID-19 term and a negative result term
  - Test type, date, and outside location terms are output if present
    - Current requirement to be added to the CSDR: [COVID term] + [negative term] + [exact date]

- Date types
  - Exact date ("4/13/2021")
  - Partial date ("4/13")
  - Relative date ("yesterday")

NLP system architecture of the NegCov pipeline
Negative COVID-19 Tests

• Identification of non-VA negative COVID-19 tests
  • Sensitivity: 97.5%

• Identification of a negative COVID-19 test with date
  • Precision: 90.5%
  • Sensitivity: 94.3%

• Limitations
  • Extracted date unrelated to COVID-19 test
  • Hypothetical test statement (“We will schedule procedure after patient tests negative for covid”)

Purpose of CSDR

• Goal:
  • Provide value as **quickly** and **consistently** as possible

• Process:

```
Collaboration    Transparency    Continual improvement
```
More Info, Help, Questions

• CSDR Wiki: COVID-19:Shared Data Resource - VA Phenomics Library

• Additional Trainings and Cyberseminars:
  • VINCI Central
  • HSR&D Cyberseminars
    • Introduction to NLP
    • Introduction to the COVID-19 Shared Data Resource
    • NLP and the CSDR – more detail on the ventilation phenotype

• Helpdesk: VINCI@va.gov