COVID-19 Evidence Accelerator Collaborative

Lab Meeting # 42

Thursday, October 7th, 2021, 3 - 4:00 pm ET

Call Summary

Overview of Lab Meeting 42

Lab Meeting 42 featured three presentations on using various real-world data (RWD) sources to support pandemic response. First, we heard from Dr. Evgeniy Gabrilovich & Tomer Shekel of Google Health about their use of anonymized Google Search data for near real-time monitoring of the spread of COVID-19. Second, Oscar Wahltinez and Philip Nelson shared how they are aggregating data from sources around the world to create a large repository of COVID-19 epidemiologic and vaccine data that can support decision-making. Finally, Dr. Donald Chalfin and Dr. Vivek Singh of Siemens Healthineers described their development of a new artificial intelligence (AI) based algorithm that can be used by hospitals to predict a patient's risk for developing severe COVID-19.

Supporting Syndromic Surveillance and COVID-19 Forecasting with Symptom Search Trends

Dr. Evgeniy Gabrilovich & Tomer Shekel, Public & Environmental Health Team, Google Health

COVID & Population Health Surveillance

- Early on capacity to test for COVID-19 infection was limited → delays in test results → limited ability to detect and respond to new outbreaks
- Stay at home orders → restricted access to medical facilities → decreased monitoring of persons with chronic disease & decreased ability to identify new cases of chronic disease
- **Need:** New methods/approaches for monitoring population health

Google Search Data

- Aggregated and anonymized search data could compliment traditional monitoring methods
    - Large cohort with broad demographic coverage
    - Broad yet granular geographic coverage
    - Near real-time data
    - Low-cost data collection
- Search data help to reveal those experiencing symptoms, those seeing a doctor/ getting tested, those visiting urgent care/ER, those admitted to a hospital

COVID-19 Symptom Search Dataset

- Aiming to augment traditional public health data sources
- Public dataset showing anonymized search trends overtime for 400+ health symptoms, signs, and conditions (2017 to present)

- Data updated daily at the county level (or at equivalent geo-granularity)
- Data available for the United States, United Kingdom, Australia, New Zealand, Ireland, and Singapore
- Key research & product opportunities:
    1. Real-time early warning systems and improved disease-spread models using COVID-19 symptom searches
    2. Identifying secondary impacts of COVID-19 (e.g., increase in depression) to catalyze action
    3. Enabling large-scale epidemiology research (COVID-related or in other disease areas) by external researchers

Findings

- Searches for Ageusia & Anosmia closely correlated with COVID-19 case counts and hospitalizations. Trends reflected outbreak intensity across different states.

Search Data for COVID Forecasting: Google Cloud AI Research

- Symptoms searches (cough, anosmia, fever, chest pain, infection, chills, shortness of breath) among top-ranked features for the fitted compartmental model rates, especially for hospitalization rate.
- Ablation studies show that for the US State-level model, symptoms search features yield ~13% MAE improvement for confirmed forecasts and ~4% MAE improvement for death forecasts
- Used by Harvard GHI to optimize testing allocations, US states to inform social distancing measures and re-opening decisions, US DoD to allocate resources, etc.
- Used by Carnegie Mellon (Delphi group) to predict (at the hospital referral region level) whether there will be a relative rise in cases by 25% in the next 7 days.

<u>COVID-19 Open Data</u>

Oscar Wahltinez, Google

Google's Response to COVID-19

- Exposure notifications
- Surfacing authoritative information in search
- Google.org: $100 million + technical expertise
- Epidemiological modeling toolkit

Epidemiology Modeling Toolkit

- Set of resources made available by Google to equip modeler and policy makers with the necessary tools to make informed decisions on how to best address the pandemic.
- Provides: Public aggregations of external data, evaluation and release of Google datasets, and publication/open sourcing of models, simulators, techniques

Key Tenets

1. **Easy to use** – easily accessible, open-source data, enables researchers/users to explore & use data
2. **Automated** – able to be maintained in the long term
3. **Transparent –** clear communication about where data comes from and what is done to it
4. **Reproducible** – open-source, should be able to be cloned by other researchers

What is COVID Open Data

- One of the largest repositories of COVID-19 epidemiological data with a powerful set of covariates in one place.
- Unique in the way it merges daily time-series, multiple international sources, at a fine spatial resolution, using a consistent set of region identifiers.
    - Epidemiological data – covid cases and deaths, good granularity globally (most areas have data at county/ equal municipality-level)
    - Vaccination data – expanding coverage of these data globally, not as granular as epidemiologic data
- Easy access to relevant data from over 20,000 distinct global locations
- Empowers users to build data visualizations, mathematical models, AI/ML training, etc.
- What's Inside?
    - Collection of datasets structure using star data schema, indexed by time & location.
    - 8 data categories: vaccinations (deployment, access, insights), epidemiology, hospitalizations, demographics, mobility reports, google search trends, government interventions (Oxford, Law Atlas), economy, geography, weather
        - Epidemiological data, hospitalizations, and demographics data are stratified by age and sex

Use of COVID-19 Open Data

- Used by researchers & data scientists for a range of informative analyses
- Used by policymakers & "decision makers" to make impactful public policy:
    - Planning safe returns to work, masking/ social distancing/ vaccination policy planning
    - To support community-based testing initiatives

<u>The Atellica COVID-19 Severity Algorithm</u>

Donald Chalfin, MD, MD, MPH & Vivek Singh, PhD, Siemens Healthineers

AI's Role Across the COVID-19 Patient Pathway

- **Diagnosis** – diagnostic testing to determine whether a patient has COVID-19
- **Prognosis & Therapy** – lab testing, hematology/hemostasis labs, blood glucose tests, CT, C-arm, X-ray, ultrasounds, to determine how severe a patients' infection is and how it should be treated
- **Follow-up** – CT, X-ray, C-arm, Ultrasound, Immuno/chemistry/hematology lab tests, molecular lab tests, to determine whether a patient is recovered from COVID-19

COVID Challenges AI Can Help Solve

- How to prioritize allocation of previous ICU beds?

- How do you determine which patients are at highest risk of severe illness and complications?

Scientific Approach to Create the Algorithm

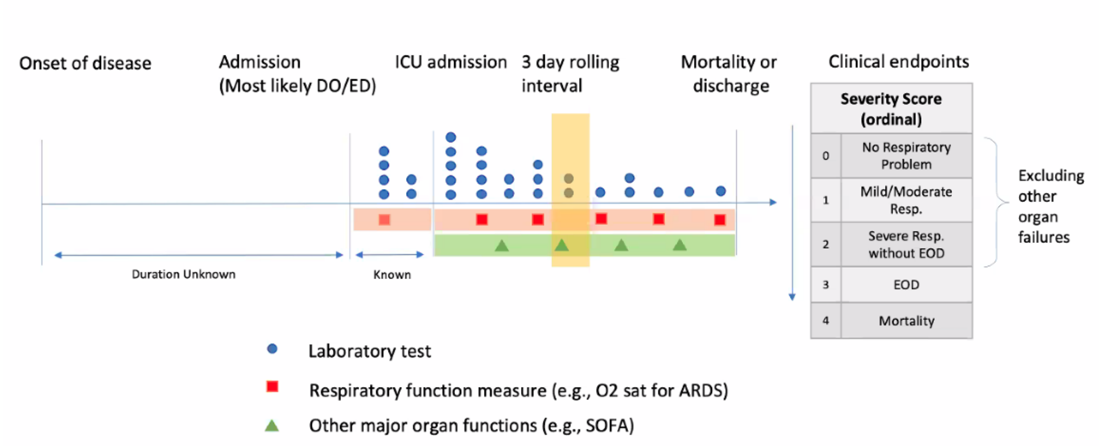
- COVID-19 Severity
    - Uncontrolled inflammation (cytokine storm) causes severe manifestation of COVID-19
    - Severity is further exacerbated by various clinical comorbidities
    - Lab markers capturing immune response and various inflammation and coagulation processes could be early predictors of severity
- Multi-site Collaboration
    - Establishing outcome-matched data-focused collaborations is key for developing an algorithm based on retrospective datasets
    - Clinical utility and usefulness of a CDS predictive system needs to be established
    - An agile prospective study is required to create evidence require to establish clinical utility
- AI Predictive Algorithm
    - Leveraging AI Factory to perform analysis using an updated version of Deep Profilter
    - Achieved good results for predicting mortality (AUC > .84) based on multi-site test cohort of 4302 cases
    - Inroads to developing Multi-Analyte Algorithmic Assays for other clinical indications

Atellica COVID-19 Severity Algorithm

- Based on our knowledge of the biology of COVID-19/ key processes associated with the disease, developed an algorithm designed to support prediction of which COVID-19 positive patients may experience severe illness
- Collaborated with healthcare institutions in Atlanta, Houston, New York, and Madrid to collect and analyze de-identified patient data
- Created AI-based algorithm designed to predict likelihood of acute respiratory failure, end organ failure, in-hospital 30-day mortality
- Algorithm trained using combined data source of ~14,500 cases
- Algorithm considers patient age and 9 laboratory measured biomarkers (D-dimer, LDH, Lymph%, Eos%, creatinine, CRP, FERRITIN, INR, Troponin-I all generated within first three days of hospitalization)
    - Investigational use only – evaluations to establish clinical utility are in progress

Data Collection Timeline

- Data collection begins when patient presents at ER with symptoms/positive test and continues until mortality or discharge from the hospital
- Based on data collected (lab tests, respiratory function measures, other organ function measures), case is assigned a severity score 0-4



AI-Based Severity Prediction Algorithm

- Challenges
  - Combination of categorical and continuous data sources
  - Reference ranges unknown
  - Implicit (sub) population-based data imputation might be required
  - Higher order (non-linear) relationship among data sources in a multi-variate setting making decision boundary to separate various sub-populations complex
- Solutions
  - Combinatorial selection of groups of data sources and performing ablation studies
  - Explicit data normalization parametrization folded into the training
  - Implicit data projection and imputation through extensive use of VAE (trained through auto-reconstruction loss with stochastic connections)
  - Perform classification and regression on the lower dimensional latent space (distilled down version of the input data)
  - Bootstrapping to est. confidence intervals for performance of cross-validation experiments
- Training/Testing Model
  - Full Model -- 57 lab measurements (within 3 days of submission), demographics (age, BMI), 6 comorbidities
  - Parsimonious Model – 10 parameters, 9 lab markers + age
  - Baseline Model – only considered IL6
  - Filtered out patients with <4 lab markers taken <3 days of admission
  - No overlap between training and testing datasets (date ranges different for each)
  - AUC = .84 +- .01 for detecting severe cases leading to mortality

Sub-group Analysis Mortality TTE from Admission

- 69 Lab Parameters + 11 Comorbidities
  - Prediction accuracy highest for mortality events closest to the admission
  - Prediction accuracy holds up even for mortality events third week after the admission
- 9 Lab Parameters + Age

    - Prediction accuracy similar to the full model for mortality events closest to admission
    - Prediction accuracy degrades for mortality events second week after the admission (may be due to other confounding factors)

Next Steps for Atellica's COVID-19 Severity Algorithm

1. **Completion of the Investigational Use Only evaluation at La Paz University** Hospital – In progress, blinded observational prospective study with the predictive model automatically generated for COVID-19 patients as part of the test panel ordered by physicians and directly integrated within workflow.
2. **Quantitative Performance Analysis** – assessment and comparison of the predictive model's actual performance compared to design and evaluation of the perceived clinical value impact
3. **Determination to Pursue Clinical Use** – If warranted, complete project activities to make algorithm available for use as a Class I, exempt non-device CDS based on US regulatory guidance.

Opportunities for Collaboration

a) De-identified Data for Additional Studies – seeking additional sources of de-identified data including confirmed outcomes for COVID-19 patients to investigate additional scenarios.

b) Evaluation Partners – Looking for additional sites to participate in retrospective studies to help evaluate performance of the predictive model.