



## **COVID-19 Evidence Accelerator Collaborative**

### **Lab Meeting #13**

*Thursday, July 16, 2020, 3:00-4:00 pm ET*

#### **Call Summary**

##### **Introduction to Lab Meeting 13**

The theme of this week's lab meeting was "COVID-19: And then what...?" Specifically, this lab meeting explored the persistent health effects of COVID-19 post-recovery and the methods the real-world evidence (RWE) community can use to learn more about the long-term health effects of COVID-19 after recovery.

The meeting began with an overview of the findings from a research letter published in JAMA detailing the experiences of 143 people who had recovered from severe COVID-19 infection who, at 60 days post-recovery, were still experiencing lingering symptoms.<sup>1</sup> This research letter underscored that there is a story after initial diagnosis and resolution of COVID-19, which needs to be examined by the community as we continue to think about disease natural history and the development of therapeutics for COVID-19.

Five presentations were given:

1. Presentation on The Need to Know More of the Story (Jacqueline Corrigan-Curay, FDA)
2. Presentation on OHDSI Community Efforts on COVID-19 Disease Natural History: Status Update and Look Forward to 'Life After COVID' (Patrick Ryan, Janssen Research and Development, Columbia University)
3. Presentation on Using Twitter to Characterize Drug Use and to Detect Symptoms/Conditions of COVID-19 (Juan Banda, Georgia State University; Daniel Prieto-Alhambra, University of Oxford)
4. Presentation on Measuring COVID-19 and Influenza in the Real World via Person-Generated Health Data (Luca Foschini, Evidation Health)
5. Tokenization to Preserve Patient Privacy in COVID-19 Research (Jason LaBonte, Datavant)

##### **Lab Meeting Presentations**

###### **The Need to Know More of the Story**

- Initially, the community prioritized identifying risk factors for COVID-19, what happens when patients get the disease, and what happens when patients are hospitalized with COVID-19. It is

---

<sup>1</sup> <https://jamanetwork.com/journals/jama/fullarticle/2768351>

increasingly becoming apparent, however, that the community must also focus on the downstream health consequences of COVID-19 and longitudinal COVID-like disease issues.

- As new therapeutics are developed to treat COVID-19 patients, these longitudinal effects will be important to consider.
- The RWE community is uniquely positioned to address many of the issues related to the longitudinal health of these patients.
- There are three key issues with our current understanding of the disease:
  - To date, most studies have focused on the acute management of COVID-19
  - Resolution of disease is often considered “the end: of the story, but it often is not!
  - The disease hasn’t been around long enough to provide a window into longitudinal effects
- With these issues in mind, the FDA has outlined a few key “needs”:
  - To develop an understanding of the natural history “post-COVID-19”
  - For information to be available to inform:
    - Clinical trial design
    - Healthcare delivery
    - Clinical management
    - Communication with patients

#### **Discussion Insights**

Post-COVID-19 data will be critical to informing trial design both in terms of efficacy and outcomes (modeling) and safety monitoring (depending on the expected toxicities of the drug, do patients need additional safety monitoring while on study).

#### **OHDSI Community Efforts on COVID-19 Disease Natural History: Status Update and Look Forward to ‘Life After COVID’**

- OHDSI promotes open collaboration which requires full transparency in every step of the research process. All details of their work can be found online, and their studies are considered a living evidence repository where any data partners can execute analyses and share aggregate results at any point.
- Through their CHARBYDIS study, OHDSI has attempted to come up with a systematic approach to studying and characterizing COVID-19 disease natural history. To do so, they created a series of target cohorts representing the life of patients at different stages of disease:
  - Persons tested for SARS-COV-2
  - Persons tested positive for SARS-COV-2
  - Persons with COVID-19 diagnosis record OR a SARS-COV-2 positive test
  - Persons hospitalized with positive test for SARS-COV-2
  - Persons hospitalized with COVID-19 diagnosis record or a SARS-COV-2 positive test
  - Persons hospitalized and requiring intensive services with positive test for SARS-COV-2
  - Persons hospitalized and requiring intensive services with COVID-19 diagnosis record OR a SARS-COV-2 positive test

- OHDSI is interested in understanding what happens to a patient before they enter a cohort, at the moment in time when they enter a cohort, and after a patient exits a cohort. This moment in time of entry into a cohort is considered the index date.
  - OHDSI is considering adding a cohort for people post-discharge after COVID-19 diagnosis record or a SARS-COV-2 positive test.
- In addition to separating patients into these cohorts, OHDSI stratifies on a number of factors including age, gender, race, index month, and certain comorbid conditions.
- The CHARYBDIS study draws data for a patient from before the index date to summarize medical history.
- At this time, patients are followed for up to 30 days post index date, but as the community continues to think about understanding life after COVID-19 recovery, more thought should be given to how long patients should continue to be followed after index.
  - Additionally, the community will need to identify which measures are important for understanding patient status post COVID-19 recovery.
- The CHARBYDIS study includes a multitude of data partners representing a diversity of data sources. By standardizing data from these sources, but not making it homogenous, OHDSI is able to embrace data heterogeneity and to explore the effects of this heterogeneity.
  - The heterogeneity allows OHDSI to explore the reliability of the phenotypes they have created.
- Thus far, OHDSI has observed that symptoms are often inconsistently captured across data sources. This could be because there is not a requirement to capture symptoms for billing purposes.

#### **Discussion Insights**

The research community needs to embrace rather than reject data heterogeneity by exploring as many different sources as possible, asking the same question of these data sets, and learning from the different answers.

#### **Using Twitter to Characterize Drug Use and to Detect Symptoms/Conditions of COVID-19**

- Twitter has been gaining attention for health-related research since 2009. Already in 2020 there have been over 600 studies published and indexed on PubMed using Twitter as the data source.
- There are a number of benefits of using Twitter for research purposes:
  - Good population representation
  - Everybody can post and have an account
  - Anonymity = unfiltered opinions
  - Data is freely available (to a certain extent)
  - Tons of data generated each day (hundreds of millions of tweets)
- Traditionally, there are also a number of disadvantages associated with using Twitter for research purposes:
  - Messy data (plenty of misspellings, shorthand, emojis, etc.)—there are at least 25 different ways people misspell hydroxychloroquine!

- Attribution is an issue—are people just mentioning something or did it happen to them?
- Freely available data is only a 1% sample of the whole set.
- Collection is hard and needs to be ongoing for days/weeks before getting considerable mass.
- COVID provides an opportunity to use Twitter to generate highly focused data:
  - Thus far, there are over 490 million tweets
  - Only COVID-related chatter is included
  - Twitter is a longitudinal data source
- Currently, work is being done to use Twitter to characterize the mentions of COVID-related drug ingredients such as hydroxychloroquine and remdesivir. Thus far, the characterization across countries is “CHARYBDIS-like.”
- Work is also being done to detect symptom/conditions related to COVID-19.
  - Symptoms/conditions such as pneumonia, infection, influenza, cough, anxiety, etc. have been cited.
- Ongoing work is being done to use this symptom and drug information to find people who had COVID-19 and their symptoms after infection.

### **Measuring COVID-19 and Influenza in the Real World via Person-Generated Health Data**

- Evidation Health is a company that uses person-generated health data (PGHD) to answer research questions of interest to biopharma and health care companies.
  - PGHD shouldn't be thought as a static dataset, as it enables the ability to recontact, follow up, and verify, allowing for continuous evidence generation.
- Evidation launched a study with the goal of understanding the utility of PGHD for early detection, monitoring, and management of COVID-19 in everyday life.
- The objectives of the study were to:
  - Quantify prevalence and progression of COVID-19 symptoms both inside and outside the clinic, and contextualize with those from other ILIs (influenza)
  - Understand behavioral and physiological parameters measured from wearable sensors
- Evidation had previously developed a survey for a participatory influenza-like illness surveillance program run on the Achievement platform ([www.myachievement.com](http://www.myachievement.com)), a 4-million-participant, US-wide virtual research site.
  - This program was amended on March 30, 2020 to include COVID-19 related items.
- Two types of data were considered in this analysis:
  - Person-reported data: SARS-CoV-2/influenza test, symptoms, health characteristics
  - Passive sensor data: Total sleep time, resting heart rate, daily step count
- Results from this study indicate that symptoms for COVID-19 cases tend to last longer than symptoms for influenza cases with similar demographics and other baseline characteristics. Notably, fatigue is reported by a larger fraction of the COVID-19 patients and for a longer duration of time than for influenza patients.
- Other preliminary results indicate that the steps lost due to COVID-19 persist longer than for flu and unlike it happens for flu, complete return to pre-illness baseline was not observed during the study period.
- Evidation is continuing to evaluate the data from this effort with a study end date of August 2020.

- Data will be shared as a part of a consortium of biometric data for COVID-19 on Sage Bionetworks Synapse Platform.
- Evidation is also embarking on a number of new initiatives including:
  - COVID-19 Pulse 200,000+ person nationwide initiative tracking people’s health and attitudes during the pandemic
  - COVID-19 early symptom detection prospective study (with BARDA, Bill & Melinda Gates foundation)
  - COVID-19 Experience Study (with NYC Health, Mt. Sinai)
- The medRxiv preprint of this study can be found at: <https://www.medrxiv.org/content/10.1101/2020.05.28.20115964v1>
- Lab meeting notes from Evidation’s previous presentation describing this study can be found at: [https://evidenceaccelerator.org/sites/default/files/2020-06/COVID-19 Evidence Accelerator Lab Meeting 8.pdf](https://evidenceaccelerator.org/sites/default/files/2020-06/COVID-19%20Evidence%20Accelerator%20Lab%20Meeting%208.pdf)

### **Tokenization to Preserve Patient Privacy in COVID-19 Research**

- Health data is fragmented and difficult to share in identified form, which poses challenges to the research community when looking past the acute setting (where data is captured in a single repository like a hospital EHR)
- One method for resolving this issue is to de-identify data by removing protected health information (PHI) while utilizing encrypted tokens to preserve linking.
  - Create encrypted key (“token”) from name, DOB, and gender or other identifying inputs (address, SSN, etc)
- Tokens are:
  - Irreversible— hashing destroys the input personally identifiable information (PII); tokens are de-identified under HIPAA
  - Unique— each token is unique to the input PII, avoiding false matches
  - Consistent— tokens created in any data set are compatible and can be matched
- Multiple tokens derived from different PII elements can be added to each patient’s record to enable accurate matching (avoiding false positives and limiting false negatives).
- Tokens allow:
  - Tracking of patients across providers
  - De-duplication of pooled records
  - Matching data to be pulled from external sources
- Tokens can be used to aggregate the RWD of COVID-19 patients wherever they seek care to better understand the downstream sequelae of COVID-19
- Tokens can also be added to clinical trial cohorts, allowing researchers to follow those patients using RWD for longitudinal follow up and monitoring (while preserving the benefits of the trial design like randomization)