COVID-19 Evidence Accelerator Collaborative

Lab Meeting #9

Thursday, June 11, 2020, 3:00-4:00 pm ET

Call Summary

Introduction to Lab Meeting 9

This week's lab meeting began with a call to action by former Food & Drug Administration (FDA) Commissioner Dr. Andrew C. von Eschenbach. Dr. von Eschenbach implored members of the Accelerator to "appreciate both the forest and the trees" when seeking to use real-world evidence (RWE) to answer COVID-19 questions.

Following Dr. von Eschenbach's call to action, Dr. Amy Abernethy noted two important actions the RWE community needs to take as it continues to explore and expand the use of RWE in COVID-19:

- To make sense of and contextualize current COVID-19 RWE results
- To build on the learnings from these results by posing additional directions for exploration to the community

Dr. Abernethy then introduced this week's lab meeting presentations:

1. Presentation on The View from Sentinel: Overview of Activities and Natural History (Michael Nguyen, FDA; Sarah Dutcher, FDA)
2. Presentation on The National COVID-19 Cohort Collaborative (N3C) (Joni Rutter, NCATS; Ken Gersing, NCATS DCI; Sam Michael, NCATS CIO)
3. Presentation on Syntegra Synthetic Data Sets (Michael D. Lesh, Syntegra.io; Ofer Mendelevitch, Syntegra.io)

Lab Meeting Presentations

Presentation on The View from Sentinel: Overview of Activities and Natural History

- Broadly speaking, the FDA's Sentinel System utilizes real-world data sources by matching them to fit its regulatory needs.
- The Sentinel Data Initiative uses several data sources that include HCA Healthcare, TriNetX, and other Sentinel Data Partners.
- The Sentinel System uses these RWD sources to address three main regulatory needs:
    - Drug use monitoring for drug shortage
        - HCA Healthcare and other Sentinel Data Partners provide sequential drug monitoring capabilities
        - TriNetX provides simple rapid querying capability

    - COVID natural history cohorts
        - HCA Healthcare
        - TriNetX
        - Natural history of coagulopathy in COVID-19
        - Feasibility assessment of integrated data systems and PCORnet
    - Methods
        - Validation of claims-based COVID algorithm
        - Master protocol for COVID natural history
- COVID-19 Natural History Cohorts in HCA Healthcare
    - The objective of the Sentinel System's Natural History work is to explore various claims-based definitions to identify hospitalized COVID-19 patients.
    - This project will describe the following among patients hospitalized with COVID-19:
        - Characteristics
        - Conditions
        - Complications
        - Treatments
        - Outcomes
    - Conditions, symptoms, complications, and outcomes are identified using diagnosis and procedure codes.
    - Medication administrations are identified using a combination of codes and text searches.
- Several preliminary findings are emerging from the Sentinel System's natural history work:
    - A higher proportion of patients hospitalized with COVID-19 present with pulmonary complications on admission than present with pulmonary symptoms after hospital admission
    - Almost 90% of patients receive an antibiotic (mostly azithromycin) during their COVID-19 hospitalization.
    - Almost 55% of patients received hydroxychloroquine, and many patients received hydroxychloroquine in combination with azithromycin.
- Further detail on the Sentinel System's coronavirus activities can be found at: https://www.sentinelinitiative.org/drugs/fda-sentinel-system-coronavirus-covid-19-activities

Discussion Insights

The Sentinel System and the Evidence Accelerator are complimentary initiatives which can draw learnings from one another for the advancement of RWD for COVID-19. Thus far, these learnings have included the application of standard definitions for COVID-19 patients, the design of studies using RWD, and the consideration of the impact of multiple types of data on RWD study findings.

Presentation on The National COVID-19 Cohort Collaborative (N3C)

- N3C includes a network of over 60 academic centers across the United States. In partnership with their network, N3C will be able to rapidly collect and analyze clinical, laboratory and diagnostic data from hospitals and health care plans. This data will be aggregated and harmonized and will be accessible for collaborative analytics and FAIR (Findable, Accessible, Interoperable, and Reusable) sharing.
- Since its launch in mid-April, N3C has received 70 requests for data access, had 38 data transfer agreements signed, completed 16 IRBs, and sent data to 7 sites.
- Although data governance and security are always essential when working with patient data, N3C recognizes the heightened level of importance associated with having a centralized data store. The N3C governance workstream includes:
    - Data transfer agreement—enables sending limited data set to NCATS
    - Governance MOP—central IRB, data access committee
    - Data use agreement—how data can be used, attribution, ensuring FAIR principles
- N3C has a dual-purpose workstream for phenotype and data acquisition:
    - Work with the community to write and maintain a computable phenotype for COVID-19
    - Write and maintain a series of scripts to execute the computable phenotype in each of four common data models: OMOP, i2b2/ACT, PCORnet, and TriNetX
- Through public/private partnerships, N3C has embarked on a data sharing initiative which uses two data types: synthetic data algorithmically-derived from limited datasets and real patient data. Preliminary comparison between computer-derived synthetic data with real patient data demonstrated the capabilities of synthetic data when testing a sepsis prediction model, which showed a high degree of similarity between the two data types.

Discussion Insights

The N3C real-world data store is a critical RWD resource which has been generated in way that is usable by numerous community members. In addition to making data accessible, it is important that researchers harmonize phenotype definitions and analysis plans so that results can be shared in a meaningful way.

Presentation on Syntegra.io Synthetic Data Sets

- Syntegra.io uses large real-world patient-level datasets to produce synthetic patient-level data in order to minimize barriers to data-sharing while maintaining patient privacy.
    - Sharing RWD with third parties will be critical for reaping the benefits of precision medicine and accelerate innovation for COVID and beyond.
- These synthetic datasets are statistically identical to the original dataset and are valid for any kind of statistical analysis, as well as training machine learning models. Synthetic datasets can help increase the percentage of patients with rare conditions, normalize bias, and can remove access friction and data use limitations.
- The Syntegra platform begins with de-identified healthcare data from EMRs, clinical trials, and claims and, using a series of steps including a pattern encoder and synthetic data generator, produces synthetic data equivalents. The pattern encoder, which sits within the data provider's security zone, is a very large deep neural network (NN) that effectively encodes the relations

between all features in the original data. Only the NN weights are passed over the firewall where a random number generator samples from the very high dimensional probability distribution to produce any number of synthetic equivalents.

- Synthetic data generated for two datasets demonstrate that univariate distributions, pair-wise correlations, and predictive models are consistent between the synthetic data and real data.
- Validation studies have tested both statistical accuracy between real and synthetic data, maintenance of privacy, and absence of memorization.
- The method is fundamentally different than other technologies intended for a similar purpose, such as differential privacy, homomorphic encryption or rule-based synthetic patient simulation.
- In partnership with the Bill & Melinda Gates Foundation, Syntegra will be using its platform for a COVID-19 project with the goal of enabling rapid sharing of sensitive COVID-19 patient-level data.
- Syntegra will use U.S. and non-U.S. COVID-19 RWD and clinical trial data to produce synthetic RWE of HCQ-treated individuals from multiple global sources.

Discussion Insights

The Syntegra platform intends to accelerate access to RWD by creating statistically identical patient-level synthetic data. RWD is a statistical sample of a larger population; synthetic data should reproduce that same statistical distribution and be subject to the same limitations. It will be important to build trust of synthetic data, as a representation of RWD, through an open and transparent framework, and validation studies. For example, government agencies and universities could make de-identified patient-level data from previously completed clinical trials available for benchmarking. Third parties would determine whether synthetic equivalents of the real data produce the same clinical trial results as the real data itself.