COVID-19 Evidence Accelerator Collaborative

Lab Meeting #11

Thursday, June 25, 2020, 3:00-4:00 pm ET

Call Summary

Introduction to Lab Meeting 11

This week's lab meeting was focused on informing the development of principles and metrics to assess the trustworthiness of real-world data (RWD) results and evidence. The Accelerator provides a "safe space" for key players across the ecosystem to lead, scrutinize and "get this right" and help develop RWD priorities. Presentations in this lab meeting demonstrate examples of scaled RWE capabilities for answering COVID-19 questions while ensuring trustworthy results:

1. Presentation on Initial Efforts Towards the Characterization of the PCORnet COVID-19 Population (Keith Marsolo, Duke University School of Medicine)
2. Presentation on 5 Countries, 96 Hospitals to Characterize COVID-19 Course in 4 Weeks (Isaac "Zak" Kohane, Harvard University)

Lab Meeting Presentations

Presentation on Initial Efforts Towards the Characterization of the PCORnet COVID-19 Population

- PCORnet is a "network of networks" that harnesses the power of partnerships between Clinical Research Networks (CRNs), Health Plan Research Networks (HPRNs), Patient Partners, and a Coordinating Center.
- PCORnet operates as a distributed network: data stays local and is transferred into a common data model (CDM) which standardizes the data into a single language, enabling fast insights.
- PCORnet leverages a number of real-world data sources and, therefore, must ensure that the underlying data are suitable for research. To do so, PCORnet utilizes a two-stage data curation process:
    - Foundational: Transform raw data into research-ready data
    - Study-specific: Ensure data are fit-for-purpose for a given study or analysis
- As of March 2020, PCORnet is conducting close to 1500 quality check measures on their data each time it is refreshed.
- PCORnet took into account a number of factors when deciding to use its network to support infectious disease surveillance and research:
    - All the core data elements existed to support COVID-19 research and surveillance and additional data elements needed have a home within the PCORnet CDM.
    - The current expectation within PCORnet was that partners refresh their CDM every quarter and run a comprehensive data quality assessment.

        - This means that by the time data has been refreshed and quality assessments have been completed, data are approximately one to three months old.
        - Under normal circumstances this timeline is appropriate, but during a pandemic it poses additional challenges.
- With these factors in mind, PCORnet began its initial COVID-19 response with the goal of characterizing the cohort of COVID-19 patients and providing detailed information on demographics and pre-existing conditions.
- To do so, PCORnet created a rapidly refreshed stand-alone version of the CDM that includes coronavirus patients plus other patients with respiratory illnesses since January 2020.
    - This process included the identification of key data elements such as COVID-19 + indicator, SARS-CoV-2 test results, and remdesivir use, some of which needed to be added because they had not existed before the COVID-19 pandemic.
    - Guidance to facilitate the loading of new data elements was also developed.
- In addition, PCORnet has utilized a "wave approach" by developing a query that is reissued weekly so sites will have an opportunity to join the effort once they are ready.
- Over the course of 2 months, PCORnet issued 7 queries, with the number of respondents growing to more than 40 sites.
- Next steps for the PCORnet COVID-19 work include:
    - Transition for surveillance to CER
        - Proposed initial use cases: coagulopathy & MIS-C
        - Data validation: provenance survey, chart review, sensitivity/specificity analyses
        - Push COVID-related elements into full CDM
    - Develop relationships with other stakeholders to leverage infrastructure
        - CDC, FDA, Reagan-Udall Foundation, FOCR

Presentation on 5 Countries, 96 Hospitals to Characterize COVID-19 Course in 4 Weeks

- Clinically relevant information to inform treatment and care decisions is lagging. Intelligence regarding "facts on the ground" is necessary for clinical decision-making to find answers to questions to help clinicians better plan (e.g. are patients with cancer at higher risk of COVID-19? Do ACE inhibitors improve prognosis in COVID-19 hypertensive patients?)
- The i2b2 tranSMART Foundation was established with the following goals:
    - To move quickly: Early intelligence is worth more than complete intelligence later
    - To leverage the i2b2/TranSMART community
    - To avoid short term tasks that require regulatory dispensation
    - To share data at the aggregate level
    - To rapidly refine the cycle of evidence generation and analysis
- Initial learnings came out 4 weeks later and are posted on CovidClinical.net and MedRxiv, which have been critically important in this effort to share data rapidly without stopping progress while peer-review is underway.
- The initial focus of the group was on comorbidities and temporal changes in key laboratory test values
- Some of the conclusions indicated a lot of variability across all of the hospitals in the ways in which different labs were reporting values (e.g. different units or measurements being used) and the importance for recognizing these differences.

- Initial findings (all available on the website) from this initiative demonstrated that:
    - There is early evidence of major pathologies: systemic inflammation, coagulopathy
    - There was greater between-hospital variation for laboratory test performance than between country variation. Despite differences in mortality across countries, patients being admitted into the hospital across countries are similar. In fact, countries are more similar than sites within them.
- The analysis of aggregated local/site level data provides a surprising opportunity, but there is no substitute for a "boots on the ground" approach. It was valuable and necessary to be able to connect directly back to the data aggregators.

Discussion Insights

Principles Needed to Generate Trustworthy Results Using RWD

- **Trust but verify:** We need the ability to verify the data to ensure credibility and accountability
- **Traceability:** Data needs to be traceable back to the data source to understand nuances of how the data was collected and processed
    - It is important to understand why the data was created, why the entity has access to the data, and where in the data supply chain the data came from
- **Usability:** The usability of the data needs to be evaluated to understand if it is fit-for-purpose to answer specific research questions
- **Source of Variability:** Need an understanding of whether variability in the data is due to data quality or the delivery of care within a health system
- **Inclusion of Data Generators:** Need to include the individuals who generate the data in the analytic phase
    - Spending time with source institutions to understand things such as the work put into ensuring the fidelity of mapping and the components that went into the common data model may be necessary.
- **Data Sharing:** The data being analyzed needs to be shared to everyone in the network to enable multiple checks before the final analysis
- **Standards for Data Transformation**: A set of standards for what goes into the data transformations

Opportunities to Foster Trust in RWD

- Often, researchers using RWD do not know the mappings for their dataset and must trust that they have been done appropriately.
    - Pre-processing efforts can be fit-for-purpose and working with those who generate the data can alleviate concerns.
- Trust is built over time and through the formation of relationships, not with code or by sharing code.
    - We can share processes, curation approaches, and comparisons.
- We should not miss this opportunity to include patient/public centered principles for trustworthiness. To earn this trust, researchers must ensure legitimacy, credibility, and accountability.
    - Suggestions for earning this trust include:

        - Commitment to ongoing, continuous, and shared learning with patients/public
        - Open communication and accountability
        - Transparency about the data source
        - The sharing of findings in a meaningful and understandable way
- Trust in data does not necessarily mean "no error" in the data but rather, an understanding of the nature of the errors in the data.
- The field is in need of guidance, a set of rules, or a checklist to certify RWD trustworthiness. Researchers who do not adhere to these rules should be held accountable.
- We need to demonstrate to sites the value of coding to standards.
    - If the source data is coded more and more to regulated data standards, the focus may shift from trusting mappers to tapping regulated data feeds from the source.

Encountered Challenges

- It can be difficult to get information about certain data set characteristics due to confidentiality concerns and because much data is sold through IT vendors that are distanced from the healthcare setting.
- In the age of the politicization of science, it may be difficult to control the "fit-for-purpose" use of data.
- Sometimes data do not follow or track to observational data principles because they are open cohort mixes of retrospective and prospective patients. This is a challenge for "fit-for-purpose" use.
- It is important to understand how representative EHR data is of reality and clinical practice. It is easy to lose this connection by centralizing and standardizing data outside of health systems.
- Not knowing what happened at every transformation step means we may have original data that is not represented by our transformed data.