# COVID-19 Evidence Accelerator Collaborative

## Lab Meeting #19

*Thurs., September 17, 2020, 3 - 4:00 pm ET*

Call Summary

**Introduction to Lab Meeting 19**

This week's lab meeting consisted of one long presentation on observational data and causal inference, and two shorter discussions both on collection and uses of real-world data amid COVID-19:

- **Causal Inference from Observational Data: Emulating a Target Trial –** *Miguel Hernán, Harvard School of Public Health*
- **COVID Patient Data Registry –** *Aneesh Chopra, Care Journey; Bala Hota & Alan Simmons, Rush University; Daniel Kurowski, Health Care Cost Institute*
- **Large Scale Data Visualization: From Rapid Insights to RWE Generation –** *Jennifer Stacy, TriNetX*

The meeting wrapped up with the "Data Visualization of the Week," a new feature of Lab Meetings. This week's visualization, provided by the Health Care Cost Institute, highlighted the impact of COVID-19 on certain preventive care services.

**Causal Inference from Observational Data: Emulating a Target Trial**

*Miguel Hernán, Harvard School of Public Health*

- The standard scientific approach to estimating causal effect – such as answering questions about comparative effectiveness and safety – is to conduct a randomized trial.
- Sometimes randomized trials are expensive, unethical, impractical, or untimely.
    - In these situations, we can analyze observational data to make a causal inference.
    - Misconception that lack of randomization in observational analyses cannot support causal inference.
- It helps to use a "Target Trial," or hypothetical randomized trial, to guide observational analysis of causal questions.
    - If the causal question cannot be translated by the Target Trial, the question is not well defined.
    - Algorithm for Causal Inference:
        1. Ask a causal question by specifying protocol of the Target Trial
        2. Answer the causal question
            a. Conduct Target Trial
            b. Use observational data to **explicitly** emulate the Target Trial and apply appropriate causal inference analytics
- Biases in observational analysis arise when there are deviations from the Target Trial protocol.
    - Ex. 1 – Postmenopausal hormone therapy and heart disease

        - Observational epidemiologic studies **compared current users vs. nonusers** to find current users of hormone therapy had >30% lower risk of heart disease.
        - Randomized trial **compared initiators vs. non-initiators** to find initiators of hormone therapy had >20% risk of heart disease.
        - Problem here is that the Target Trial was not explicitly emulated and follow-up time was initiated years after initiation of hormonal therapy:
            - Observational study – current (prevalent) users vs. nonusers
                - Prevalent users depleted of susceptibles → current use became a marker of not being susceptible
            - Randomized trial (Target Trial) – initiators (incident users) vs. non-initiators
        - The problem here is not lack of randomization, rather the question of the Target Trial was not explicitly emulated.
    - Ex. 2 – Statins and Cancer
        - Observational studies reported an association between statins and lower cancer risk (some studies found an implausible 50-65% lower risk). However subsequent meta-analyses of randomized trials show no effect.
        - When emulating the Target Trial, results mirrors those observed in randomized trials. Assessing previous observational studies several issues were discovered:
            - Included prevalent users at baseline
            - Using postbaseline information (observed duration of statin therapy) to assign baseline treatment status (leading to immortal time bias)
- In both examples, the problem with the observational studies was **not** lack of randomization. Lack of randomization still presents issues but many issues in observational studies or more fundamental:
    - Explicitly emulating the Target Trial can eliminate "self-inflicted" biases such as selection bias, immortal time bias, etc. Incorrect specification of time zero is often the actual culprit.
    - Confounding is not self-inflicted and can still arise even when explicitly emulating a target trial.
- 2 Key Components:
    1. Randomized assignment – emulation requires adjustment for confounding.
    2. Specification of time zero – time 0 must be synchronized with determination of eligibility and assignment of treatment strategies.
- Added complication – journals have an aversion to the use of causal language in observational studies.
    - They do not let you say that your goal is causal – rather they require you to speak of your study question in terms of "association"
    - Need to be able to explicitly describe causal goal – not being able to is like shooting without a target.

**<u>COVID Patient Data Registry</u>**

*Aneesh Chopra, Care Journey*

- Robert Wood Johnson Foundation Data COVID-19 Sharing Collaborative aimed to bringing together databases to answer pressing COVID questions.

*Bala Hota & Alan Simmons, Rush University*

- Using observational & real-world data to answer questions requires organization and planning. (i.e. Electronic Health Record data collection, concept mapping, staging data, creating a common data model).
- Multi-site data sharing is complex because of gaps and inconsistencies across research databases:
    - Variable levels of implementation and familiarity with common data models (OMOP)
    - Misaligned shared or common definitions (i.e. How do you define what constitutes a COVID-19 patient? What are appropriate time points to share data? How to count the number of hours on ventilation? How do you standardize definitions of race & ethnicity?)
    - Processes of data governance and legal review vary, which can slow down data sharing.
- Main challenge: How to make sure you have data updated weekly that is usable for many different data models?
    - Develop code that is shareable.
    - Set of common definitions.
    - Share "toolkit":
        - Document set of processes followed to create the database shell.
        - Document art used to make decisions to represent the information.
    - Following instructions given in the toolkit increases the likelihood there is common data.

*Daniel Kurowski, Health Care Cost Institute*

- Shared results from this effort.
- 5 Health Systems' data used – Mt. Sinai, UC, Rush, Geisinger, NWMC
- Data Visualizations were shared on the following:
    - Total Cases by Facility
    - Distribution of Cases by Facility by Week
    - Average Age of COVID Patients by Week
    - Distribution of Cases by Age Group (0-54 & 55+)
    - Distribution by Sex Overtime

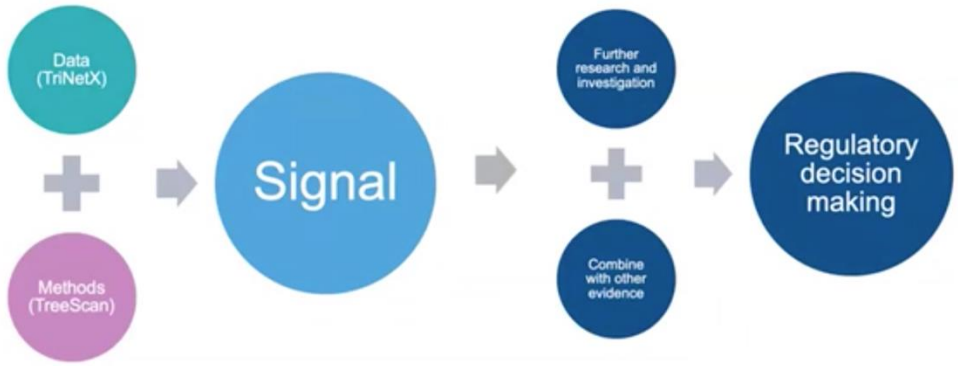
**Large Scale Data Visualization: From Rapid Insights to RWE Generation**

*Jennifer Stacey, TriNetX*

- TriNetX used inspiration from early COVID headlines & research insights to develop their focus question: Deep dive into which COVID-19 treatments are working and which are not?
- Data platform used enables data visualization through every step of research.
    - Organized, Concise, Rapid Insights
    - 95 HCOs
    - Defined COVID population = 385,000+ severe/ inpatient COVID patients
    - Used different codes for insights about different cohorts of patients
    - Able to see what classes of drugs (down to their ingredients) were being used to treat these patients.
    - Also used detailed visuals for looking at what top treatments and treatment combinations were used.
- These rapid insights lead to further research questions such as:

- Mechanical ventilation among patients treated with dexamethasone vs. not exposed to dexamethasone → mortality analysis shows dexamethasone patients have better survival

- Next Steps of Data Visualization to Scale
  - To make data more savvy and useful it helps to organize and visualize data.




  - Using signal detection visualizations can help identify research questions that need to be asked.

**Data Visualization of the Week**

*Health Care Cost Institute*

Visualization of the week

**The Impact of COVID-19 on the Use of Preventative Health Care**

Wednesday, 09 September 2020
Katie Martin, Daniel Kurowski, Phillip Given, Kevin Kennedy, Elianna Clayton

HEALTH CARE COST INSTITUTE

COVID-19 RESEARCH DATABASE

https://healthcostinstitute.org/hcci-research/the-impact-of-covid-19-on-the-use-of-preventive-health-care