COVID-19 Evidence Accelerator Collaborative

Parallel Analysis Meeting

Wednesday, July 8, 2020, 12:00-1:00 pm ET

Call Summary

Introduction

The goal of this week's COVID-19 Evidence Accelerator Parallel Analysis Meeting was to discuss the challenges encountered by the parallel analysis pioneers in addressing the first question set about hydroxychloroquine (HCQ). Solutions for addressing these challenges in the short-term as the Accelerator takes on additional question sets were proposed. Additional longer-term solutions for the real-world data community will be discussed at a later date.

Review of Previously Resolved Challenges

During a previous parallel analysis meeting, we discussed and resolved three challenges faced by the parallel analysis project in answering the first question set about HCQ:

Challenge & Context: Coding	Potential Solution: Share codes
Mechanical ventilation was challenging due to inconsistent coding.	A few of our Evidence Accelerators developed a natural language processing script to pull out this information that they are willing to share!
Case definition for COVID-19. Issues include which date to accept as first occurrence, identifying "best" criteria	• Hierarchical coding definition with lab as most specific, followed by COVID medication exposure, & presence of ICD10 code • ICD10 codes now available and code list generated & sharable
Identification of COVID-19 medication. It can be difficult to identify whether patients are receiving treatments as a part of the same administration (combination treatment) or as individual therapies	Coding algorithms to identify medications of interest as part of combination or individual treatments have been developed and can be shared.

Discussion of Additional Challenges & Solutions

Four additional challenges were identified and discussed during this Parallel Analysis meeting:

Challenge 1: Biased assessment of baseline comorbidities

Challenge & Context: Events outside the Hospital	Potential Solutions: Linkages
Biased assessment of baseline comorbidities. Claims systems & integrated care settings have membership enrollment to define a population with access to care. In non-integrated health systems, this cohort does not readily exist and biases the identification of comorbidities among those who are frequent care seekers (usually very sick and women).	• Inclusion criteria should require that in EHR data. there should be at least 1 encounter with the health system within 90-days prior to the index date.

- It can be very difficult to get information about those patients who were diagnosed and prescribed medications in the outpatient setting.
    - This lack of information poses a problem when considering person-time.
- The inclusion criteria which requires an encounter with the healthcare system is useful but may result in a substantial reduction in sample size.
- Generally speaking, in a claims environment the absence of care is meaningful. In an EHR environment, it is not meaningful because it can mean a patient did not seek care **OR** that the patient may have sought care elsewhere.
    - This is alleviated in an integrated healthcare system.
- Activity patterns are generally thought to be dependent on exposure and outcomes, a bias which is referred to as "informed presence bias" by some scholars.
- Baseline covariates should be thought of in terms of both chronic conditions and the patient's baseline COVID-related covariates.
- Those groups who have access to both outpatient and inpatient data should perform sensitivity analyses where they act as if they only have inpatient data to see if there is a difference when using both data sources and when using only inpatient data.

Proposed Solution: A set of standard sensitivity analyses to be done by data set (not by question) to assess the impact of inclusion criteria on the outcomes and conclusions should be created.

Challenge 2: How to assess outcomes outside of hospital setting

Challenge & Context	Potential Solution
How to assess outcomes outside of hospital setting. While mortality was easily identified, We recognized a need to look at composite information about a patient, such as where they were discharged to, in order to have a better understanding of a patient's outcome.	• Need to capture discharge disposition • Linkages to death data outside of hospital should be established • Claims systems can track by a member number to identify future interaction. • EHRs, other than integrated, need to tap methods used to identify re-admission. Potential solutions: ?

- Outcome measures other than mortality which could serve as a proxy or indicator for the survival outcome could be considered secondarily.
- The distinction between hospital re-admissions and re-admissions caused by re-infection should be considered.

Proposed Solution: The Datavant data set (comprised of obituary information and a death master file) and the NCI mortality tracker could be linked to for research purposes.

Challenge 3: Immortal time bias

Challenge & Context	Potential Solution
Immortal Time Bias. Given an index date=hospital admission: • In identifying COVID-19 if DX is accepted after index date. • In identifying treatment exposure • Will be an issue if doing comparative analysis.	For questions regarding treatment effect or risk, index could be treatment start, so as not to occur after start of index. Requires matching to non-exposed on similar characteristics (e.g. baseline comorbidity).

- In addition to thinking through the impact of immortal time bias on comparative analyses, we must consider the impact of this bias on side-by-side descriptive comparisons of treated and untreated populations.
    - The index date for each dataset should be described.
- The treatment start date is likely the best index date approach for our analyses, but ultimately the appropriate method depends on the question of interest.
    - Ultimately, we need to think through how best to find the counterfactual exposure and how to find comparators who have equipoise.

Proposed Solution: The treatment start date should generally be used as the index date, but the implications of this selection should be further examined for each question of interest.

Challenge 4: Time-varying confounding by indication

Challenge & Context	Potential Solution
Time-Varying Confounding by Indication. Treatments are given to the sickest who are at higher risk of morbidity and mortality. • Will be an issue if doing comparative analysis.	• Time to treatment analysis to identify factors associated with worsening symptoms that necessitate treatment • Causal Methods: IPTW, marginal structural models, G-estimation

- The potential confounding by severity of disease is a form of confounding by indication.
    - Baseline severity is the most important independent predictor of patient outcomes.
    - The impact of this baseline severity on the outcomes of patients with cancer, etc. warrants investigation.
- The methods proposed above may be effective in isolating the treatment effect.
- The impact of time-varying confounding by indication is often not linear because the sickest patients are often transitioned to hospice care.
- This challenge will depend on the dataset, the variables that are available, how time is calculated, and whether the dataset includes claims data, EHR data, or both.
- The solution to this challenge will require more specific thinking and discussion.

Proposed Solution: TBD, warrants further discussion