Characterizing Respiratory Symptoms and COVID-19 Trends from OMOP-CDM database for Public Health Reporting

Marcello Chang, MS[1*], Jonathan Lu, MS[1*], Birju Patel, MD, MPH[1], Nigam H. Shah, MBBS, PhD[1], Jonathan H. Chen, MD, PhD[1]

[1]Center for Biomedical Informatics Research, Stanford School of Medicine, Stanford, CA

* indicates equal contribution

Introduction

While healthcare systems have focused on overall case reporting to inform the COVID-19 response, finer information on patient demographics, respiratory illness, tests and results is also important. For example, population shifts in respiratory symptoms for patients not yet tested for COVID-19 may be difficult to identify with currently mandated data. The CDC does not typically have access to joint trends in patient characteristics, as these require a robust and interoperable health information infrastructure. Our objective was to demonstrate an example of public health analytics for COVID-19 by using our clinical data warehouse to provide a synthesis of demographic and clinical data from our healthcare system.

Methods

We queried the Stanford Medicine STARR-OMOP[1] de-identified data warehouse, which conformed to the OMOP-CDM schema[2] and can be accessed via Google BigQuery. We composed SQL queries for three phenotypes: 1) visits related to respiratory illnesses identified by related concept families of ICD10CM codes, 2) SARS-CoV2 Nucleic Acid Amplification tests identified by LOINC codes, and 3) results associated with those SARS-CoV2 tests. To characterize the relationship between testing and symptoms, we additionally filtered SARS-CoV2 tests that resulted up to 14 days after respiratory diagnoses. The data were then stratified by age, gender, race, and ethnicity. Metadata conveying processing parameters were also included in the final export.

Results and Lessons Learned

From 6/10/2020 to 7/21/2020, we supplied 5 data updates, cumulatively reporting on 44,240 patients with respiratory symptoms and 79,200 patients receiving SARS-CoV2 tests. We were able to extract clinical data and create a public health update in less than three minutes, which could be re-run on-demand. Example reports are available publicly at tinyurl.com/stanfordcovidcdcmmwr

There were several factors contributing to our success, including the availability of de-identified STARR-OMOP data for prototyping prior to establishing a data transfer agreement with public health agencies and the use of a common data model enabled rapid design iteration as the health questions evolved. The use of a standardized schema allows our analysis pipeline to be transferred to other institutions to enhance public health reporting. Doing so would be a major step ahead in addressing the data woes surrounding COVID19 response[3]. Limitations of the data include missing values in certain fields, which may reflect workflow issues or limitations of the data transformation for our analytic warehouse.

Conclusion

Data from the OMOP-CDM can facilitate serving an urgent public health need. Compared to traditional analytics, which often involve tailored queries, a common data model can facilitate rapid and iterative reporting for clinical trends during a pandemic.

References

1. Datta S, Posada J, Olson G, Li W, O'Reilly C, Balraj, D, Mesterhazy J, Pallas J, Desai P, Shah NH. A new paradigm for accelerating clinical data science at Stanford Medicine. Arxiv preprint.
2. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong ICK, Rijnbeek PR, Lei J van der, Pratt N, Norén GK, Li Y-C, Stang PE, Madigan D, and Ryan PB. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform. 2015; 216: 574–578.
3. Maxmen, Amy. Why the United States is having a coronavirus data crisis. Nature 2020, doi:10.1038/d41586-020-02478-z.