Title: Computable Phenotype and Severity Outcomes Validation of Adult Patients Admitted to the Hospital with Confirmed Coronavirus Disease 2019 (COVID-19)

Authors: Lauren Heery, BS[1*]; Will Carter, MIS[2]; Kristine Erlandson, MD, MS[1]; Lisa Schilling, MD, MSPH[3]

1. University of Colorado School of Medicine, Department of Medicine, Division of Infectious Diseases 2. University of Colorado School of Medicine, Adult & Child Consortium for Health Outcomes Research & Delivery Science 3. University of Colorado School of Medicine, Department of Medicine, Division of General Internal Medicine, Data Science to Patient Value Initiative

*Corresponding author; KE and LS are student's faculty mentors

Abstract: The COVID-19 pandemic presented an opportunity to apply clinical informatics advances in real-time to answer clinical questions with patient-level data. EHR-computed phenotypes of patient cohorts have been used to characterize the natural history of COVID-19, the impact of therapeutic interventions, and outcomes of healthcare systems. Despite the widespread use of these methods, it is often unknown the extent to which this data is valid and high quality. This study aims to validate an EHR-computed phenotype of patients hospitalized for COVID-19 with a database of manually abstracted patient charts. We hypothesize that the cohort definition phenotype will accurately identify patients hospitalized for COVID-19 at University of Colorado Hospital who were admitted for the first time March 18, 2020 to April 26, 2020. The EHR phenotype generated 438 patient encounters, while the chart review database included 415 patient records. There were 379 overlapping patient records. The EHR phenotype dataset had 32 false negatives, resulting in a 92.2% sensitivity compared to the REDCap database. After ETL updates and modifications of the phenotype to 1) remove date restrictions on the new COVID-19 ICD-10 code and 2) remove a requirement for a second diagnosis code specifying disease caused by COVID-19, this sensitivity increased to 99.1%. These high rates of sensitivity indicate that the construction of computable phenotypes is reliable even within the context of shifting use of diagnosis codes and viral tests for a new disease process. This work has important implications for future EHR-based phenotypes and observational research methods.

Author Disclosures: WC, LH, LS: nothing to disclose. KE: I have received research funding (to the University of Colorado) from Gilead Sciences, and consultant payment from Viiv Pharmaceuticals and Theratechnologies (to the University of Colorado).

COMIRB Statement: This project is covered by COMIRB Protocols 20-0690, 20-0730, and 20-2223. All three protocols received IRB review exemption as they involve secondary data. Certificates of exemption available upon request.