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

University of Colorado Anschutz Medical Campus

Lauren Heery, BS[1]; Will Carter, MIS[2] ;Kristine Erlandson, MD[3]; Lisa Schilling, MD, MSPH[4]

1. School of Medicine 2. ACCORDS 3. Division of Infectious Disease 4. Division of General Internal Medicine, Data Science to Patient Value Initiative

Background

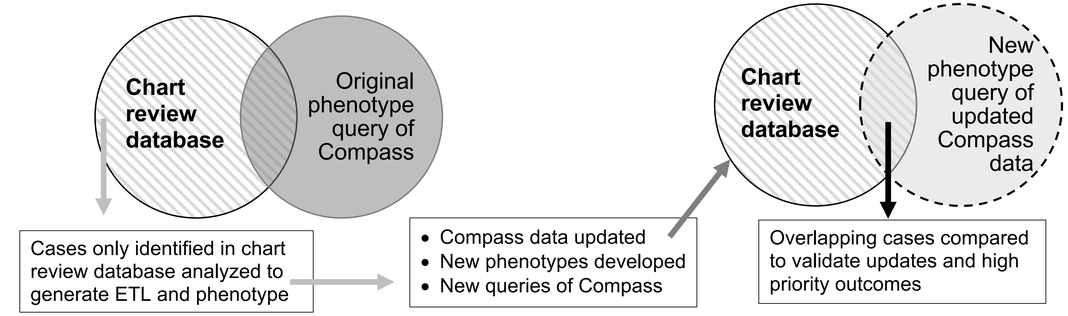
- Observational electronic health record (EHR)-derived data has been increasingly used to characterize patient-level aspects of the COVID-19 pandemic including disease natural history, the impact of therapies, and population-level outcomes (1).
- Computable phenotypes have emerged as an important informatics method (2) to solve such research questions.
- Despite the widespread use of these methods, it is unknown the extent to which this data is valid and high quality (3).

Rationale: This study aims to validate an EHR-computed phenotype of patients hospitalized for COVID-19 with a database of manually abstracted patient charts.
Hypothesis: We hypothesize that the cohort definition phenotype will accurately identify patients hospitalized for COVID-19.

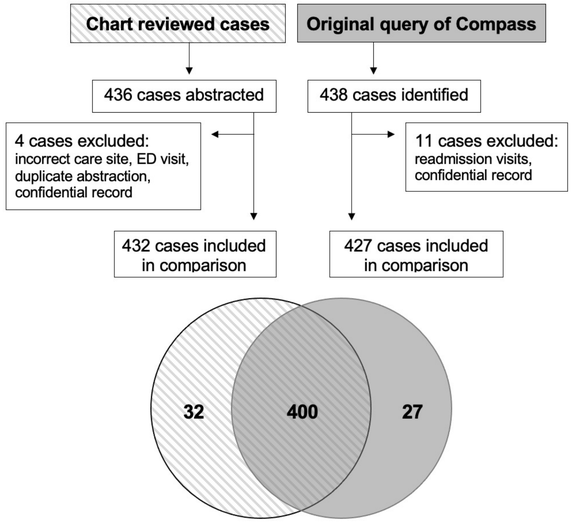
Methods

- **REDCap chart review database** of patients hospitalized for COVID-19 March 18-April 26, 2020 at UCHealth University of Colorado Hospital; inclusion based on COVID-19 infection flag in Epic, first admission
- The **EHR-computed phenotype** was developed through the Rapid Response Data for Discoveries (R2D2) collaboration and based on the Common Data Model (4)
- A query was designed to match the REDCap database cohort and search data in Health Data Compass (HDC) (5)
- Clinical and demographic variables were abstracted and extracted from flowsheets and notes in Epic charts
- Patients in the HDC dataset were linked by an arbitrary identifier with the patients in the REDCap database
- Patient records were manually compared to determine validity of the phenotype and outcomes of interest

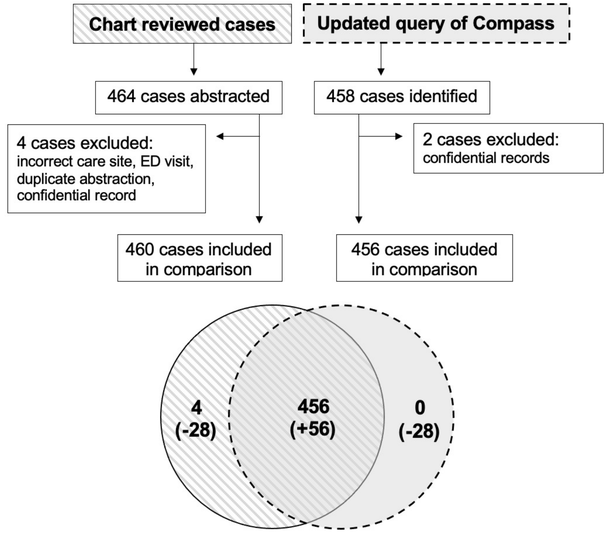
Methods : Validation Process Overview



Results : Initial comparison between chart reviewed cases and Compass data query



Results : Repeat comparison between chart review cases and updated Compass data query



Conclusions

- An EHR-computed phenotype had a 92.2% sensitivity compared with chart review in identifying cases of patients hospitalized for COVID-19.
- Analysis of patients missed by the EHR phenotype identified errors in data extraction, transformation and loading (ETL), particularly with regard to visit type in patients' encounters.
- This work highlighted limitations of applying date-restricted logic in the phenotype definition to ICD-10 diagnosis codes.
- Analysis of outcomes of interest (invasive mechanical ventilation, extracorporeal membrane oxygenation, and in-hospital mortality) showed high quality concordance.
- Demographic information reliably underwent ETL from the EHR into HDC.

Implications and Future Work

- This work has led to improvements in the phenotype definition's inclusion criteria and Compass's process.
- Data quality validation is critical as EHR-derived data and phenotypes are increasingly used in observational research.
- These findings ensure integrity of this EHR-data and strengthens the reliability of the Common Data Model.

References 1. Haendel MA, Chute CG, Bennett TD, et al. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. J Am Med Inform Assoc. 2021;28(3):427-443. doi:10.1093/jamia/ocaa196 2. Richesson RL, Hammond WE, Nahm M, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. J Am Med Inform Assoc. 2013;20(e2):e226-231. 3.Callahan TJ, Bauck AE, Bertoch D, et al. A Comparison of Data Quality Assessment Checks in Six Data Sharing Networks. EGEMS (Wash DC). 2017;5(1):8. Published 2017 Jun 12. doi:10.5334/egems.223 4. Boyce RD, Ryan PB, Norén GN, et al. Bridging islands of information to establish an integrated knowledge base of drugs and health outcomes of interest. Drug Saf. 2014;37(8):557-567. doi:10.1007/s40264-014-0189-0 5. Kim J, Neumann L, Paul P, et al. Privacy-Protecting, Reliable Response Data Discovery Using COVID-19 Patient Observations. Preprint. medRxiv. 2020;2020.09.21.20196220. Published 2020 Sep 23. doi:10.1101/2020.09.21.20196220

Acknowledgements Thanks to REDCap chart review team: Lyndsey Babcock, Connor Fling, Kellen Hirsch, Dave Sheneman, Nemanja Vukovic, and Taylor Wand. Thanks to Health Data Compass team: Ufi Olakpe, Michelle Edelmann, Michael Kahn, and Ian Brooks. Supported by NIH/NCATS Colorado CTSA Grant Number TL1 TR002533, NHLBI GEMS Program Grant Number 5R25HL103286-10 and 5R25HL103286-12. Contents are the authors' sole responsibility and do not necessarily represent official NIH views. Supported by Health Data Compass Data Warehouse project (healthdatacompass.org). REDCap supported by NIH/NCATS Colorado CTSA Grant Number UL1 TR002535. Contents are the authors' sole responsibility and do not necessarily represent official NIH views. The Rapid Response Data for Discoveries (R2D2) Collaboration is funded by the Gordon and Betty Moore Foundation.

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).