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.
- 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 27, 2020 at UCHHealth 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

Chart review database

Original phenotype query of Compass cases only included in chart review database analyzed to generate ETL and phenotype

Cases published in chart review database had a 92.2% sensitivity with regard to ETL and phenotype

Chart review database

New computed phenotype query of updated Compass database

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

<table>
<thead>
<tr>
<th>436 cases abstracted</th>
<th>436 cases identified</th>
<th>432 cases included in comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>27</td>
<td>7</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>464 cases abstracted</th>
<th>456 cases abstracted</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

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 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


Acknowledgements

2. Lauren Heery, BS; Will Carter, MS; Kristine Erlandson, MD; Lisa Schilling, MD, MSPH

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

Disclosures

Lauren Heery, BS: I have received research funding from the University of Colorado Non-Symmetric Grants, and consultant payment from Viiv Pharmaceuticals and Theratechnologies (to University of Colorado).