Validation of an internationally derived patient severity phenotype to support COVID-19 analytics from electronic health record data

Siyue Yang
PhD Student, Department of Statistical Sciences
University of Toronto
June 28, 2021

Validation of an internationally derived patient severity phenotype to support COVID-19 analytics from electronic health record data

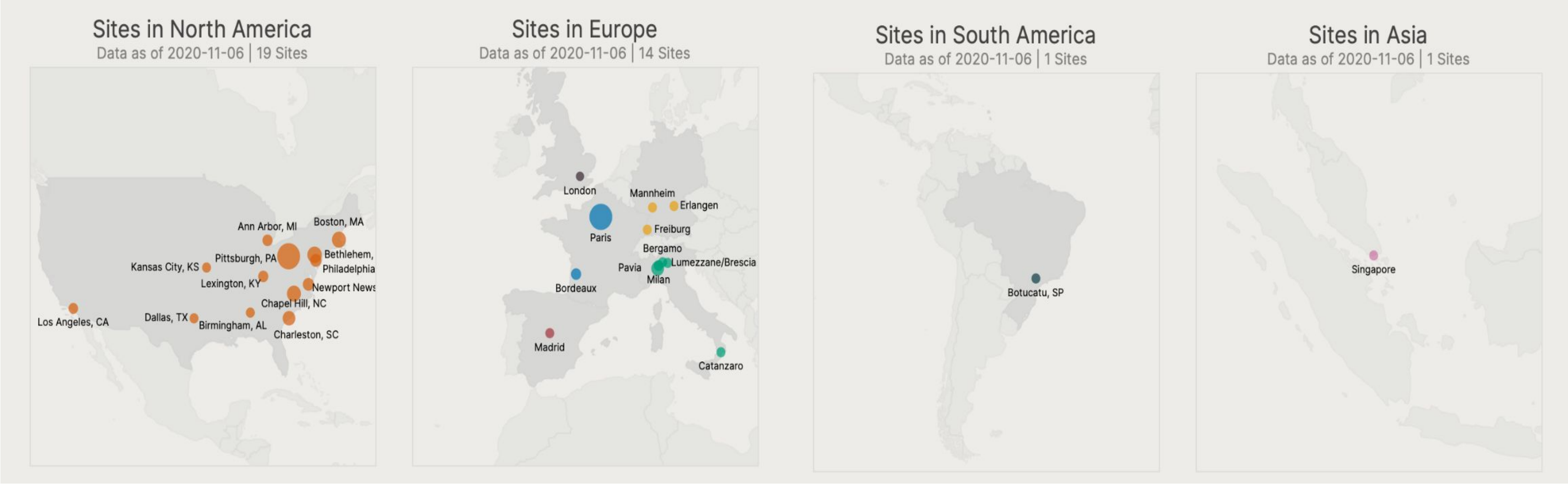
- By Jeffrey G. Klann, Hossein Estiri, Griffin M. Weber, Bertrand Moal, Paul Avillach, Chuan Hong, Amelia L.M. Tan, Brett K. Beaulieu-Jones, Victor Castro, Thomas Maulhardt, Alon Geva, Alberto Malovini, Andrew M. South, Shyam Visweswaran, Michele Morris, Malarkodi J. Samayamuthu, Gilbert S. Omenn, Kee Yuan Ngiam, Kenneth D. Mandl, Martin Boeker, Karen L. Olson, Danielle L. Mowery, Robert W. Follett, David A. Hanauer, Riccardo Bellazzi, Jason H. Moore, Ne-Hooi Will Loh, Douglas S. Bell, Kavishwar B. Wagholikar, Luca Chiovato, Valentina Tibollo, Siegbert Rieg, Anthony L.L.J. Li, Vianney Jouhet, Emily Schriver, Zongqi Xia, Meghan Hutch, Yuan Luo, Isaac S. Kohane, The Consortium for Clinical Characterization of COVID-19 by EHR (4CE), Gabriel A. Brat, and Shawn N. Murphy
- JAMIA 2021; ocab018

Validation of an internationally derived patient severity phenotype to support COVID-19 analytics from electronic health record data

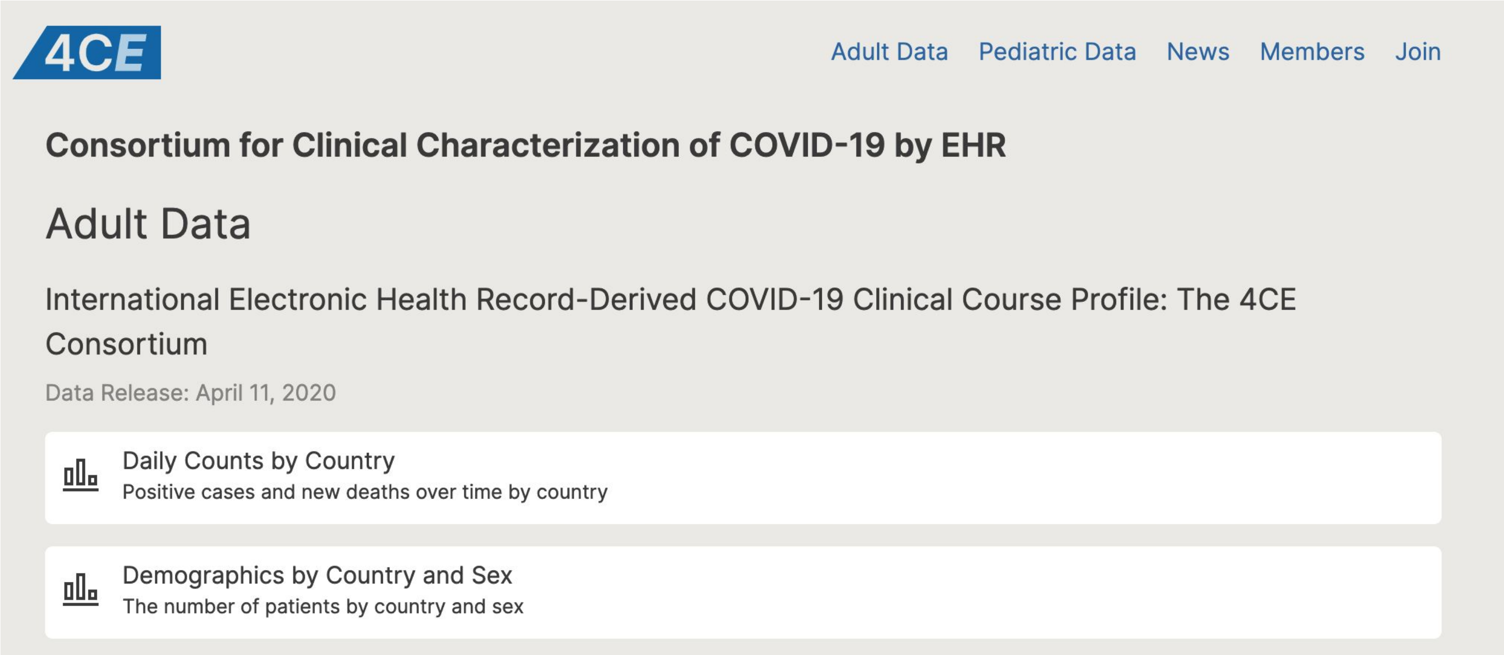
- By Jeffrey G. Klann, Hossein Estiri, Griffin M. Weber, Bertrand Moal, Paul Avillach, Chuan Hong, Amelia L.M. Tan, Brett K. Beaulieu-Jones, Victor Castro, Thomas Maulhardt, Alon Geva, Alberto Malovini, Andrew M. South, Shyam Visweswaran, Michele Morris, Malarkodi J. Samayamuthu, Gilbert S. Omenn, Kee Yuan Ngiam, Kenneth D. Mandl, Martin Boeker, Karen L. Olson, Danielle L. Mowery, Robert W. Follett, David A. Hanauer, Riccardo Bellazzi, Jason H. Moore, Ne-Hooi Will Loh, Douglas S. Bell, Kavishwar B. Wagholikar, Luca Chiovato, Valentina Tibollo, Siegbert Rieg, Anthony L.L.J. Li, Vianney Jouhet, Emily Schriver, Zongqi Xia, Meghan Hutch, Yuan Luo, Isaac S. Kohane, The Consortium for Clinical Characterization of COVID-19 by EHR (4CE), Gabriel A. Brat, and Shawn N. Murphy
- JAMIA 2021; ocab018

4CE: International consortium for EHR data-driven studies of the COVID-19 pandemic

- 35 international sites



4CE dashboards



https://covidclinical.net/data/index.html

4CE publications

International Electronic Health Record-Derived COVID-19 Clinical Course Profile: The 4CE Consortium

Data Release: April 11, 2020

Validation of a Derived International Patient Severity Phenotype to Support COVID-19 Analytics from Electronic Health Record Data

International Comparisons of Harmonized Laboratory Value Trajectories to Predict Severe COVID-19: Leveraging the 4CE Collaborative Across 342 Hospitals and 6 Countries: A Retrospective Cohort Study

Data Release: December 18, 2020

Multinational Prevalence of Neurological Phenotypes in Patients Hospitalized with COVID-19

What Every Reader Should Know About Studies Using Electronic Health Record Data but May Be Afraid to Ask

4CE publications

International Electronic Health Record-Derived COVID-19 Clinical Course Profile: The 4CE Consortium

Data Release: April 11, 2020

Validation of a Derived International Patient Severity Phenotype to Support COVID-19 Analytics from Electronic Health Record Data

International Comparisons of Harmonized Laboratory Value Trajectories to Predict Severe COVID-19: Leveraging the 4CE Collaborative Across 342 Hospitals and 6 Countries: A Retrospective Cohort Study

Data Release: December 18, 2020

Multinational Prevalence of Neurological Phenotypes in Patients Hospitalized with COVID-19

What Every Reader Should Know About Studies Using Electronic Health Record Data but May Be Afraid to Ask

4CE publications

International Electronic Health Record-Derived COVID-19 Clinical Course Profile: The 4CE Consortium

Data Release: April 11, 2020

Validation of a Derived International Patient Severity Phenotype to Support COVID-19 Analytics from Electronic Health Record Data

International Comparisons of Harmonized Laboratory Value Trajectories to Predict Severe COVID-19: Leveraging the 4CE Collaborative Across 342 Hospitals and 6 Countries: A Retrospective Cohort Study

Data Release: December 18, 2020

Multinational Prevalence of Neurological Phenotypes in Patients Hospitalized with COVID-19

What Every Reader Should Know About Studies Using Electronic Health Record Data but May Be Afraid to Ask

Paper, Slides

Goals of the paper

1. Develop and validate an **expert-driven computational phenotype for COVID-19 severity** based on hospital data from 12 international sites

2. Pilot a **data-driven phenotyping approach using machine learning** at 1 site and compare to the expert-driven approach

Study description

- **Data source**
  - 10,510 hospitalized patients with a positive COVID test from 7 days before to 14 days after hospitalization
  - Patients were from 12 sites in the US, France, Italy, Germany, and Singapore

Study description

- **Data source**
  - 10,510 hospitalized patients with a positive COVID test from 7 days before to 14 days after hospitalization
  - Patients were from 12 sites in the US, France, Italy, Germany, and Singapore
- **Analysis period**
  - Cases recorded from March to August 2020

Study description

- **Data source**
    - 10,510 hospitalized patients with a positive COVID test from 7 days before to 14 days after hospitalization
    - Patients were from 12 sites in the US, France, Italy, Germany, and Singapore
- **Analysis period**
    - Cases recorded from March to August 2020
- **Codified EHR data**
    - Demographics, diagnoses, medications, labs, procedure codes

Study description

- **Data source**
  - 10,510 hospitalized patients with a positive COVID test from 7 days before to 14 days after hospitalization
  - Patients were from 12 sites in the US, France, Italy, Germany, and Singapore
- **Analysis period**
  - Cases recorded from March to August 2020
- **Codified EHR data**
  - Demographics, diagnoses, medications, labs, procedure codes
- **Outcome**
  - COVID-19 severity

Existing COVID-19 severity definitions

- 25 prior studies investigated COVID-19 severity
- 20 used the **WHO definition** for severity
  - Fever OR suspected respiratory infection AND one of the following:
    - Respiratory rate > 30 breaths/min
    - Severe respiratory distress
    - Arterial oxygen saturation measured by pulse oximeter (SpO2) ≤ 93% while breathing room air
- 5 used **ICU admission**

Challenges with existing COVID-19 severity definitions

- 25 prior studies investigated COVID-19 severity
- 20 used the **WHO definition** for severity
  - Fever OR suspected respiratory infection AND
    - Respiratory rate > 30 breaths/min
    - Severe respiratory distress
    - Arterial oxygen saturation measured by pulse oximeter (SpO2) ≤ 93% while breathing room air

Too inclusive & can't be derived from codified data

- 5 used **ICU admission**

Not recorded at all sites

Goal 1: Expert-driven phenotype for COVID-19 severity

- Needed to obtain a definition using only the available data classes
  - Demographics
  - Diagnoses
  - Medications
  - Labs
  - Procedure codes
- Leveraged the WHO definition to **define the diagnosis group**
  - Patients who require invasive mechanical ventilation for acute respiratory failure, or vasoactive medication infusions for shocks

How to define COVID-19 severity by expert-driven approach?

- **Presence of at least one** of the following disease-related codes

	Severity definition	Standardized codes
Lab test	Partial pressure of carbon dioxide or partial pressure of oxygen (PaCO2 or PaO2)	LOINC 2019-8, LONIC 2703-7
Medication	Sedatives/anesthetics or treatment for shock	RxNorm 6130, 206967, etc
Diagnosis	ARDS, ventilator-associated pneumonia	ICD-10 J80, ICD-9 518.82, etc
Procedure	Endotracheal tube insertion or invasive mechanical ventilation	ICD-10 5A093*, 5A094*, etc

Table A1: 4CE Severity Codes

How to define COVID-19 severity by expert-driven approach?

- **Presence of at least one** of the following disease-related codes

	Severity definition	Standardized codes
Lab test	Partial pressure of carbon dioxide or partial pressure of oxygen (PaCO2 or PaO2)	LOINC 2019-8, LONIC 2703-7
Medication	Sedatives/anesthetics or treatment for shock	RxNorm 6130, 206967, etc
Diagnosis	ARDS, ventilator-associated pneumonia	ICD-10 J80, ICD-9 518.82, etc
Procedure	Endotracheal tube insertion or invasive mechanical ventilation	ICD-10 5A093*, 5A094*, etc

Table A1: 4CE Severity Codes

How to validate expert-driven definition?

- Compared derived phenotypes against gold standard

	Gold-standard Outcome	
Algorithm Phenotype	ICU and/or Death	No ICU or Death
Severe	Phenotype and outcome	Phenotype only
Not severe	Outcome only	Neither phenotype nor outcome

Table A2: Severity Analysis 2x2 tables design

How to validate expert-driven definition?

- Compared derived phenotypes against gold standard
  - Gold-standard: **ICU admission and/or Death**
    - Obtainable from EHR data
    - Really a proxy for severe disease or hospital course

	Gold-standard Outcome	
Algorithm Phenotype	ICU and/or Death	No ICU or Death
Severe	Phenotype and outcome	Phenotype only
Not severe	Outcome only	Neither phenotype nor outcome

Table A2: Severity Analysis 2x2 tables design

High variability of performance across sites

	Higher Specificity								
	GLO1	GLO2	USA5	USA8	USA1	USA3	USA6	GLO5	USA4
Sensitivity	0.35	0.74	0.58	0.66	0.76	0.75	0.73	0.83	0.67
Specificity	0.96	0.93	0.86	0.87	0.89	0.89	0.79	0.96	0.68
PPV	0.55	0.90	0.80	0.75	0.82	0.71	0.73	0.74	0.54
NPV	0.92	0.82	0.68	0.82	0.85	0.91	0.79	0.98	0.79
F-Score	0.43	0.81	0.67	0.70	0.79	0.73	0.73	0.78	0.60
F-Score CI	(0.26-0.60)	(0.74-0.88)	(0.65-0.69)	(0.68-0.73)	(0.74-0.83)	(0.55-0.91)	(0.70-0.76)	(0.67-0.90)	(0.65-0.69)

	Higher Sensitivity			
	USA7	**USA2**	GLO3	Meta-Analysis (95% CI)
Sensitivity	0.91	0.86	0.88	0.73 (0.64-0.8)]
Specificity	0.50	0.64	0.46	0.83 (0.76-0.9)]
PPV	0.70	0.70	0.63	0.73 (0.63-0.8)]
NPV	0.80	0.82	0.79	0.83 (0.75-0.9)]
F-Score	0.79	0.77	0.73	0.72 (0.63-0.80)
CI	(0.75-0.83)	(0.74-0.82)	(0.68-0.78)	

Estimates of the pooled scores were computed using a fixed-effect meta-analysis model.

4CE: Consortium for Clinical Characterization of COVID-19 by EHR; CI: confidence interval; ICU: intensive care unit; NPV: negative predictive value; PPV: positive predictive value.

Table 3: The sensitivity, specificity, PPV, NPV, and F1 score of the 4CE severity phenotype for the outcome ICU admission and/or death at each site in the United States and outside the United States (Global)

High variability of algorithm sensitivity across sites and codes

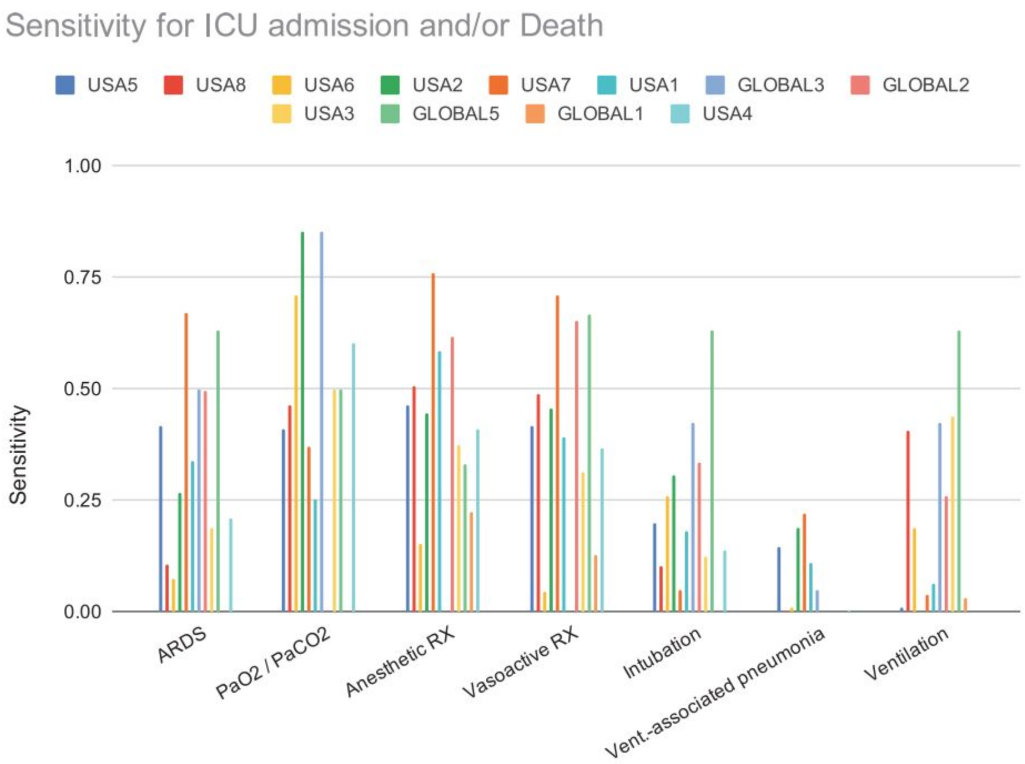


Figure 1: Sensitivity of code classes to identify intensive care unit (ICU) admission and/or death. ARDS: acute respiratory distress syndrome; PaCO2: partial pressure of carbon dioxide; PaO2: partial pressure of oxygen; RX: medication

Overlap of high-level code classes among identified severe patients

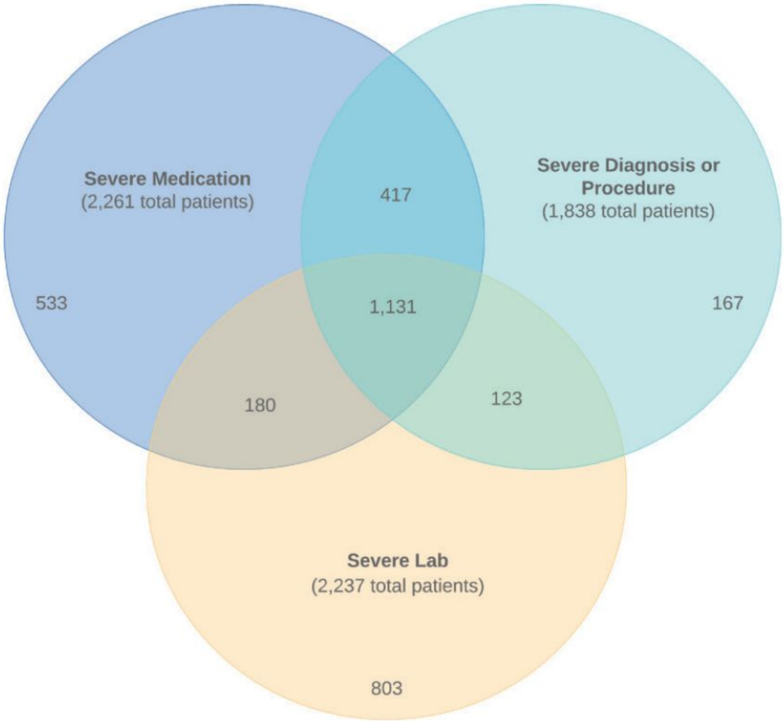
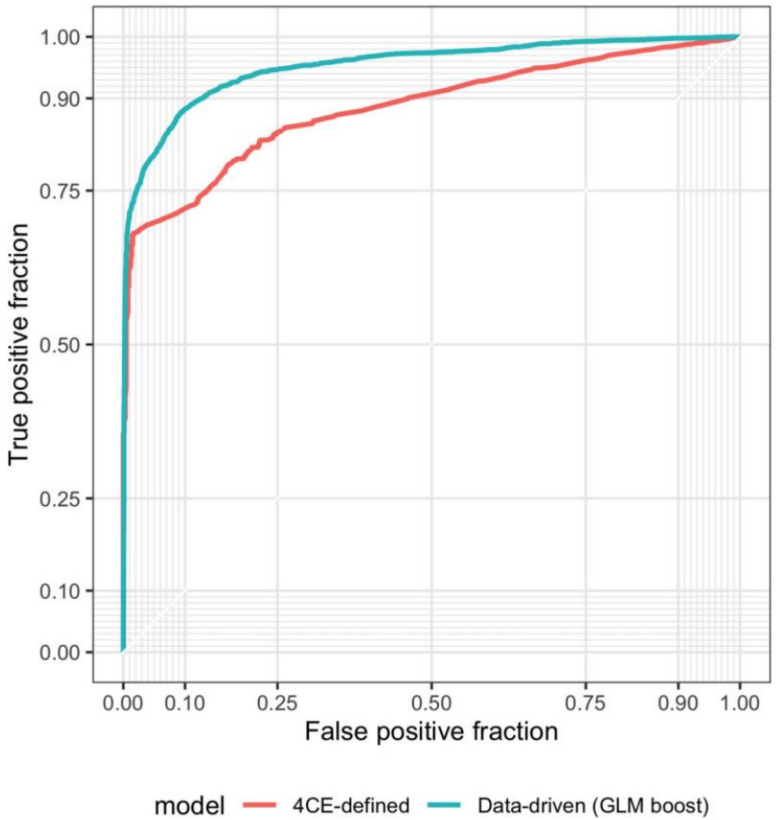


Figure 3: Venn diagram showing overlap of code classes among patients with the Consortium for Clinical Characterization of COVID-19 by EHR severity phenotype (9 sites reporting)

Goal 2: Data-driven definition of COVID-19 severity

- Developed two logistic regression models to predict ICU admission and/or death using data from 1 site
    - **Model with expert-defined features**: Codes from the 4CE expert-driven definition
    - **Model with automatically selected features**: Codes from the Minimize Sparsity Maximize Relevance method
- Compared performance with the ROC curve and AUROC

Comparison between data-driven and expert-driven approach



Model	AUROC
Expert-derived features	0.903 (95% CI, 0.886-0.921)
Automatically selected features	**0.956 (95% CI, 0.952-0.959)**

Figure 4: Receiver-operating characteristic curves when using a general linear model (GLM) boost algorithm on Consortium for Clinical Characterization of COVID-19 by EHR (4CE)–defined features vs a data-driven approach

Limitations

- Only include codified data for analysis (no clinical notes)
- Data were collected from sites during a surge in the COVID-19 pandemic
  - Could have unexpected bias in the results
- Data-driven approach was only a pilot study at 1 site

Summary

- Developed an expert-driven approach at 12 sites
    - Achieved high accuracy overall, but showed variable performance across sites
    - Required significant effort to define the disease-related codes for each site
- Piloted a data-driven approach at 1 site
    - Promising performance, but likely not generalizable
    - Hard to implement at other sites due to the chart review required for model training and evaluation

Extra slides

Details of participated sites

Healthcare System	City	Country	Hospitals	Total Beds	ICU Beds	ICU Data Source	4CE Cohort Size	Additional Codes in Value Set
Mass General Brigham (Partners HealthCare)	Boston, Massachusetts	United States	10	3418	292	Hospital data	3290	None
University of Pennsylvania	Philadelphia, Pennsylvania	United States	5	2469	515	Hospital data	2330	Hospital data for intubation and ventilation
University of Pittsburgh	Pittsburgh, Pennsylvania	United States	39	8400	589	CPT code and hospital location	990	CPT codes for intubation and ventilation
Beth Israel Deaconess Medical Center	Boston, Massachusetts	United States	1	673	77	Hospital data	690	None
University of Michigan	Ann Arbor, Michigan	United States	3	1043	141	CPT code and hospital location	420	None
University of California, Los Angeles	Los Angeles, California	United States	2	786	192	Hospital data	430	None
Bordeaux University Hospital	Bordeaux	France	3	2676	180	Hospital data	360	CCAM (French procedure codes)
Istituti Clinici Scientifici Maugeri	Pavia, Lumezzane/Brescia, Milan	Italy	3	775	0	N/A (rehab hospital—no ICU)	260	None
Medical Center, University of Freiburg	Freiburg	Germany	1	1660	132	Hospital data	190	ICD-10 GM and OPS codes
Boston Children's Hospital	Boston, Massachusetts	United States	1	404	107	ICU note type	60	None
National University Hospital	Singapore	Singapore	1	1556	65	Hospital data	260	SNOMED codes for diagnoses; TOSP billing codes for procedures
St. Luke's University Health Network	Bethlehem, Pennsylvania	United States	12	1700	287	Hospital data	1230	None

4CE: Consortium for Clinical Characterization of COVID-19 by EHR; CPT: Current Procedural Terminology; ICD-10: International Classification of Diseases–Tenth Revision–German Modification; ICU: intensive care unit; N/A: Not Applicable; OPS: Operation and Procedure Classification; SNOMED: Systematized Nomenclature of Medicine; TOSP: Table of Surgical Procedures.

Table 1. Participating 4CE sites and metadata on ICU and 4CE coding definitions, number of beds, and 4CE cohort size (rounded to the nearest 10)

Demographics characteristics

Category	Group	All Patients (N = 10 340)	Severe Phenotype Patients (n = 3800)	Severe (%)
Age	0-25 y	450 (4)	90 (3)	21
	26-49 y	2180 (21)	630 (17)	29
	50-69 y	3740 (36)	1580 (42)	42
	70-79 y	1820 (18)	800 (21)	44
	80+ y	2070 (20)	650 (17)	32
Sex	Female	4930 (47)	1610 (42)	33
	Male	5410 (52)	2190 (58)	41
Race	White	4210 (42)	1520 (41)	36
	Black	2550 (25)	1000 (27)	39
	Other	3360 (33)	1220 (33)	36

Values are n (%), unless otherwise indicated. Numbers are rounded to the nearest 10.

Table 2: Demographic characteristics of all patients vs all patients with the severity phenotype, across the 12 sites Category

Comparison between ways to obtain ICU

- Particular ICU admission coding approach could impact the measured performance

	MGH		UKFR	
	Hospital	Chart	Hospital	Chart
Sensitivity	0.58	0.80	0.74	0.85
Specificity	0.86	0.75	0.93	0.96
PPV	0.80	0.57	0.90	0.93
NPV	0.68	0.90	0.82	0.91

The hospital column is repeated from Table 2 for clarity.

4CE: Consortium for Clinical Characterization of COVID-19 by EHR; MGH: Massachusetts General Hospital; NPV: negative predictive value; PPV: positive predictive value; UKFR: University of Freiburg Medical Center in Germany.

Table 4: Comparing the performance of the 4CE severity phenotype when using chart-reviewed ICU admission data or hospital codes at MGH and UKFR

Individual outcomes in Network-wide analysis

Outcome	Measure	Meta-analysis	Mean	USA5	USA8	USA6	USA2	USA7	USA1	GLOB3	GLOB1	GLOB2	USA3	GLOB5	USA4
ICU/ DEATH	Sensitivity	0.73 [0.64, 0.82]	0.73	0.58	0.66	0.73	0.86	0.91	0.76	0.88	0.35	0.74	0.75	0.83	0.67
ICU	Sensitivity	0.77 [0.68, 0.87]	0.79	0.62	0.75	0.74	0.88	0.91	0.78	0.89	n/a	0.81	0.75	0.83	0.71
DEATH	Sensitivity	0.76 [0.64, 0.87]	0.78	0.59	0.66	0.78	0.91	0.90	0.80	0.91	0.35	0.76	1.00	1.00	0.73
ICU/ DEATH	Specificity	0.83 [0.76, 0.91]	0.79	0.86	0.87	0.79	0.64	0.50	0.89	0.46	0.96	0.93	0.89	0.96	0.68
ICU	Specificity	0.79 [0.71, 0.87]	0.75	0.85	0.85	0.78	0.62	0.45	0.88	0.41	n/a	0.89	0.89	0.96	0.67
DEATH	Specificity	0.67 [0.60, 0.75]	0.64	0.70	0.74	0.60	0.47	0.31	0.67	0.32	0.96	0.75	0.74	0.88	0.60
ICU/ DEATH	PPV	0.73 [0.63, 0.82]	0.71	0.80	0.75	0.73	0.70	0.70	0.82	0.63	0.55	0.90	0.71	0.74	0.54
ICU	PPV	0.67 [0.58, 0.77]	0.68	0.75	0.68	0.71	0.67	0.63	0.81	0.52	n/a	0.81	0.71	0.74	0.47
DEATH	PPV	0.24 [0.15, 0.33]	0.25	0.29	0.32	0.16	0.29	0.24	0.20	0.19	0.55	0.45	0.06	0.03	0.25
ICU/ DEATH	NPV	0.83 [0.75, 0.91]	0.83	0.68	0.82	0.79	0.82	0.80	0.85	0.79	0.92	0.82	0.91	0.98	0.79
ICU	NPV	0.86 [0.79, 0.94]	0.86	0.75	0.89	0.81	0.86	0.82	0.86	0.84	n/a	0.89	0.91	0.98	0.85
DEATH	NPV	0.97 [0.93, 1.02]	0.95	0.89	0.92	0.97	0.96	0.93	0.97	0.95	0.92	0.92	1.00	1.00	0.93

Table A3: Sensitivity, Specificity, PPV, and NPV by outcome (ICU admission and/or death, ICU, and death)

Comparison between data-driven and expert-driven approach

- Top 10 codes (by odds ratio) from data-driven approach
  - Conceptually similar to expert-driven approach
    - Partial pressure of carbon dioxide, partial pressure of oxygen, ARDS, sedatives
  - Reflective of **ICU ordering patterns**
    - D-dimer, immature granulocytes, albumin (labs)
  - **Proxies of severity**
    - Chlorhexidine, glycopyrrolate, palliative care encounter

Comparison between data-driven and expert-driven approach

- Data-driven has AUROC greater than expert-driven definitions

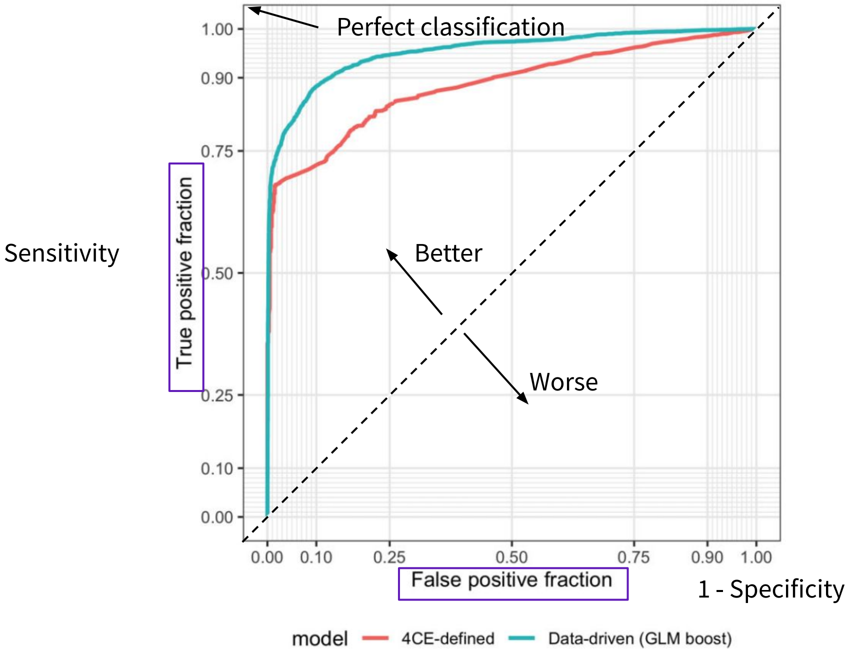
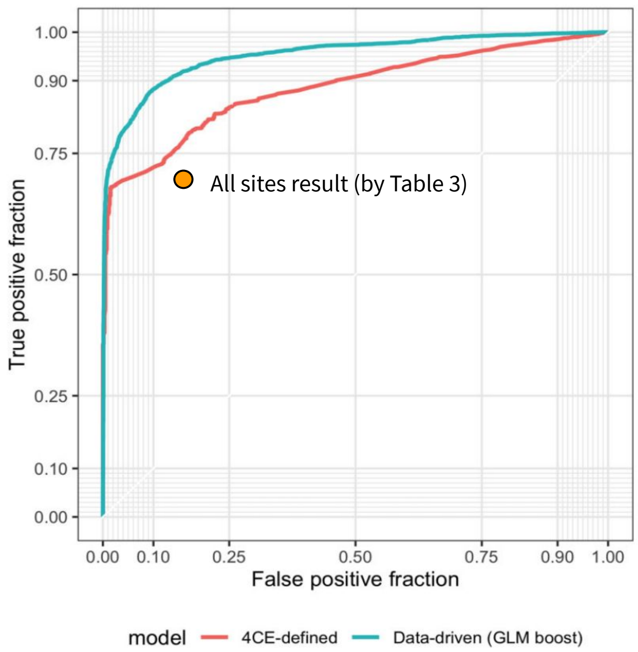
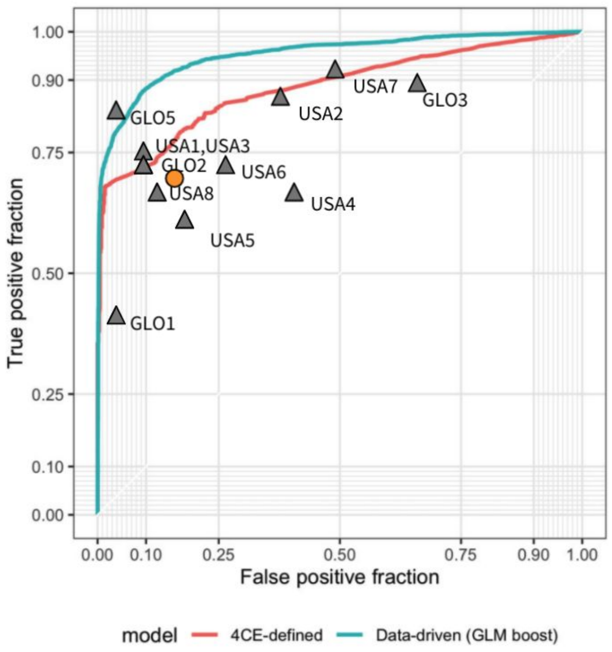


Figure 4: Receiver-operating characteristic curves when using a general linear model (GLM) boost algorithm on Consortium for Clinical Characterization of COVID-19 by EHR (4CE)–defined features vs a data-driven approach

Comparison between data-driven and expert-driven approach

- The comparison used only 1 site data - Partners HealthCare, generally have high sensitivity and specificity compared to other sites





High variability of percentage of all severe patient with a code in each code class

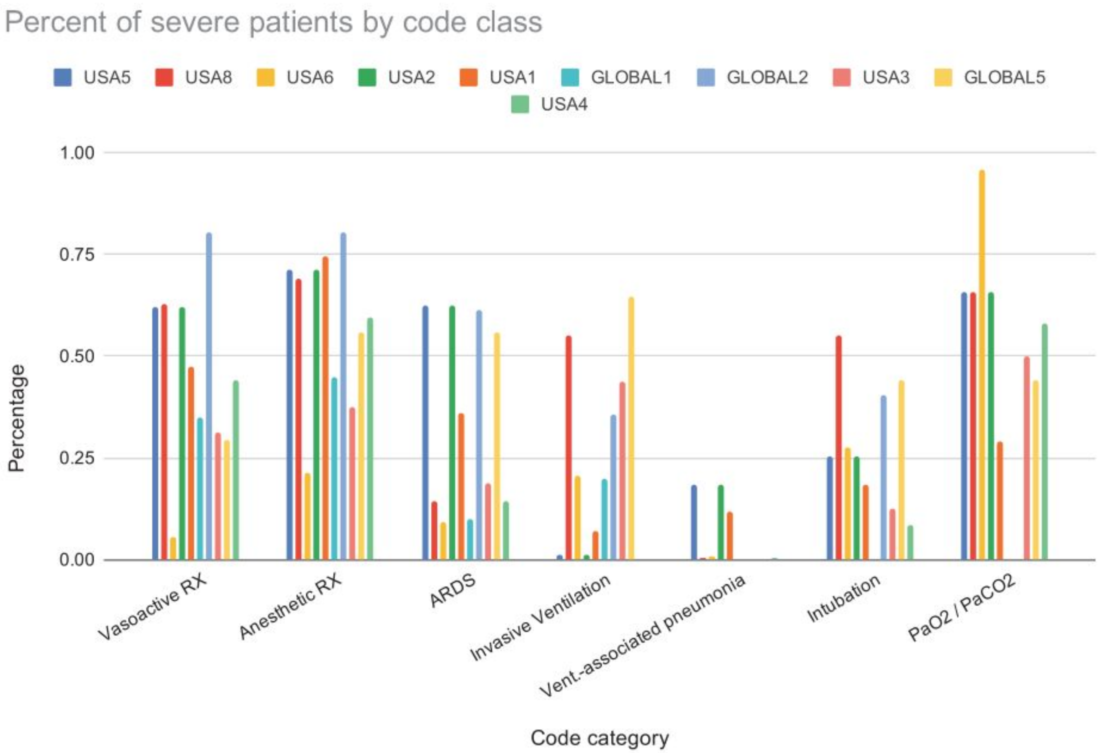


Figure 2: Percentage of patients identified by the Consortium for Clinical Characterization of COVID-19 by EHR severity phenotype, broken down by code class. ARDS: acute respiratory distress syndrome; PaCO2: partial pressure of carbon dioxide; PaO2: partial pressure of oxygen; RX: medication.

How to obtain ICU admission information?

- 3 methods for confirming ICU admission
  - Chart review
  - Local hospital data
  - Specific ICU CPT procedure codes

from most to least accurate

How to define COVID-19 severity by expert-driven approach?

- **Presence of at least one** of the following disease-related codes

	Severity definition	Standardized codes
Lab test	Partial pressure of carbon dioxide or partial pressure of oxygen (PaCO2 or PaO2)	LOINC 2019-8, LONIC 2703-7
Medication	Sedatives/anesthetics or treatment for shock	RxNorm 6130, 206967, etc
Diagnosis	ARDS, ventilator-associated pneumonia	ICD-10 J80, ICD-9 518.82, etc
Procedure	Endotracheal tube insertion or invasive mechanical ventilation	ICD-10 5A093*, 5A094*, etc

Table A1: 4CE Severity Codes