INFORMATICS PROFESSIONALS. LEADING THE WAY.

Learning Clinical Concepts for Predicting Risk of Progression to Severe COVID-19

Hit by a Pandemic: Studying the Impacts of COVID-19

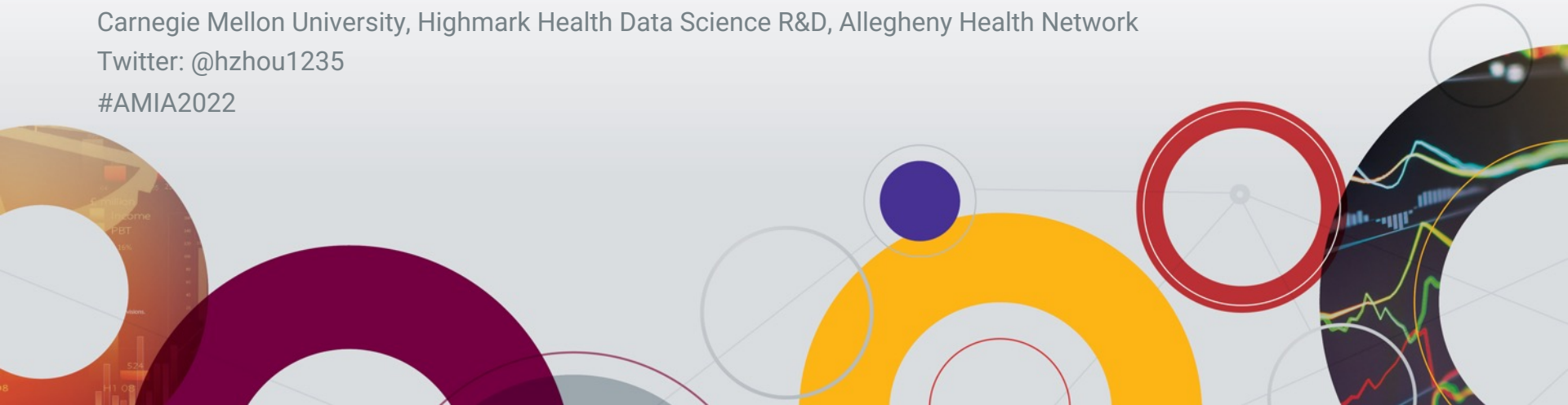
Session 27

Helen Zhou, Cheng Cheng, Kelly Shields, Gursimran Kochhar, Tariq Cheema, Zachary C. Lipton, Jeremy C. Weiss

Carnegie Mellon University, Highmark Health Data Science R&D, Allegheny Health Network

Twitter: @hzhou1235

#AMIA2022



Disclosure

I have no relevant relationships with commercial interests to disclose.

Learning Objectives

After participating in this session the learner should be better able to:

- To understand how to define positive anchors for learning clinical concepts
- To understand how to utilize these concepts for downstream tasks such as risk prediction
- To apply and analyze survival analysis models to predict risk when patient data is censored
- To evaluate models in a back-testing framework and analyze performance over time
- To understand risk factors for severe COVID-19

As COVID-19 becomes endemic…

Hospitals will continue to need to be prepared for COVID-19 patients.

Of particular interest: patients likely to progress to **severe COVID-19**

Goal: Predict COVID-19 patients' risk of progressing to severe COVID-19.

Challenges:

- Under-coded features that degrade the accuracy of smaller models
- Many features (more accurate) vs. fewer features (more aligned w/ clinical intuition)
- A constantly evolving environment (new treatments, policies, variants, etc.)

Contributions

- We develop two sets of high-performance risk scores:
  1. unconstrained risk prediction model built from **all available features**
  2. pipeline that first learns a **small set of concepts anchored to clinical intuition**
- Learned concepts:
  - **boost performance** over the corresponding features
  - demonstrate improvements over all available features when evaluated out-of-sample (in **subsequent time periods**)
- Our models outperform previous works (C-index 0.84–0.87 vs. 0.60–0.81)
- Interactive visualization tool (Sankey diagram) for understanding clinical concepts and their relation to predicted outcomes

Cohort, Outcome, Features

- Retrospective observational data collected from **Jan 2020—Jan 2022** by a major healthcare provider in Southwestern Pennsylvania
- Cohort *(n = 31,336)* of individuals **testing positive for the first time** *(t_0)*
- **Severe COVID-19** outcome (mechanical ventilation, ICU admission, or death)
  - Survival outcome (time-to-event) defined relative to t_0
- Features (also defined relative to t_0):
  - Location (inpatient vs. outpatient)
  - Demographics
  - Labs
  - Medications
  - Vaccines
  - Symptoms
  - Problem history

Cohort Characteristics

Characteristic		Count (%)	Characteristic		Count (%)
Gender	Female	17,874 (57.0%)	Location of test	Inpatient	13,246 (42.3%)
	Male	13,455 (42.9%)		Outpatient	15,868 (50.6%)
Age	Under 20	2,836 (9.1%)		Unknown	2,222 (7.1%)
	20—30	3,987 (12.7%)	Outcomes	Severe COVID-19	5,272 (16.8%)
	30—40	4,134 (13.2%)		ICU Admission	4,811 (15.4%)
	40—50	4,155 (13.3%)		Death	1,554 (5.0%)
	50—60	5,444 (17.4%)		Mechanical ventilation	1,096 (3.5%)
	60—70	5,017 (16.0%)			

Motivation for Learning Clinical Concepts

- In EHRs, potential risk factors are often recorded indirectly or unreliably
- We often observe the **presence** of a clinical condition but **not its absence** (e.g. “diabetes” ICD code → diabetes, but unmarked not necessarily non-diabetic)
- Models learned automatically often end up using **proxies** that indirectly encode important risk factors (e.g. saline IV bolus encoding inpatient status)
- One could manually map these proxies, but it is difficult to be comprehensive
- Instead, we use the *anchor-and-learn* framework (Halpern et. al. 2016) to **learn clinical concepts** corresponding to major risk factors
- These clinical concepts are then used for **downstream risk prediction**

PU Algorithm for Learning Concepts

For each (unobserved) binary concept y_c of interest, define

- **Anchor:** an observed feature conditionally independent of all other features conditioned on the concept, i.e. $p(x_c|y_c = 1) = p(x_c|y_c = 1, x_{\bar{c}})$
- **Positive anchor:** anchor x_c whose presence almost certainly implies the presence of the concept y_c (e.g. diabetes ICD code → diabetes concept)

Consider the probability of a positive anchor x_c given the other observed covariates:

$$p(x_c = 1|x_{\bar{c}})$$
$$= p(x_c = 1 \wedge y_c = 1|\, x_{\bar{c}}\,)$$
$$= p(y_c = 1|x_{\bar{c}})p(x_c = 1|y_c = 1, x_{\bar{c}})$$
$$= p(y_c = 1|x_{\bar{c}})p(x_c = 1|y_c = 1)$$
$$\Longrightarrow \boldsymbol{p(y_c = 1|x_{\bar{c}}) = p(x_c = 1|x_{\bar{c}})/\delta_c}\,, \quad \text{where } \delta_c = p(x_c = 1|y_c = 1)$$

PU Algorithm for Learning Concepts

For each binary concept y_c of interest, define

- **Anchor:** conditionally independent of all other concept, i.e. $p(x_c|y_c = 1) = p(x_c|y_c = 1, x_{\bar{c}})$
- **Positive anchor:** anchor x_c whose presence of the concept y_c (e.g. diabetes ICD code → diab

Consider the probability of a positive anchor x_c g

$$p(x_c = 1|x_{\bar{c}})$$
$$= p(x_c = 1 \wedge y_c = 1|\, x_{\bar{c}}\,)$$
$$= p(y_c = 1|x_{\bar{c}})p(x_c = 1|y_c = 1, x_{\bar{c}})$$
$$= p(y_c = 1|x_{\bar{c}})p(x_c = 1|y_c = 1)$$
$$\Longrightarrow \boldsymbol{p(y_c = 1|x_{\bar{c}}) = p(x_c = 1|x_{\bar{c}})/\delta_c},$$

where $\delta_c = p(x_c = 1|y_c = 1)$

$\Longrightarrow$ The probability of **positive vs. negative** is ***proportional*** to the probability of **positive vs. unlabeled**!

PU Algorithm for Learning Concepts

$$\Longrightarrow \boldsymbol{p(y_c = 1|x_{\bar{c}}) = p(x_c = 1|x_{\bar{c}})/\delta_c}\ , \quad \text{where } \delta_c = p(x_c = 1|y_c = 1)$$

"Anchor-and-learn" Framework: For each binary concept y_c of interest,

1. Identify some key informative observations x_c (***positive anchors***) for the concept
2. Learn a positive vs. unlabeled (PU) logistic regression classifier $g(x_{\bar{c}})$ for the probability of positive anchor given other covariates, i.e. $p(x_c = 1|x_{\bar{c}})$.
3. Estimate scaling constant $\delta_c = p(x_c = 1|y_c = 1)$ by averaging predictions on all positive examples P, that is: $\hat{\delta}_c = \frac{1}{n}\sum_{x_{\bar{c}} \in P} g(x_{\bar{c}})$
4. Scale predictions from PU classifier by constant: $p(y_c = 1|x_{\bar{c}}) = g(x_{\bar{c}})/\hat{\delta}_c$

Identifying Clinical Concepts of Interest

Clinician survey results:

1. Old age
2. Inpatient
3. Outpatient
4. Diabetes
5. Shortness of breath
6. Fever
7. Cough
8. Fatigue
9. COVID-19 vaccination
10. Flu vaccination
11. Obesity
12. Hypertension
13. Immunocompromised
14. COPD
15. Congestive heart failure
16. Chronic kidney disease
17. Hyperglycemia
18. Transplant
19. Cancer
20. Lung disease
21. Myalgia

After identifying these concepts, we identify corresponding positive anchors through string matching and clinician recommendations. Then, we proceed with PU learning.

Model

Patients are often *censored* – it's unknown what happened to them past a certain time point (e.g. discharge). Thus, we use survival analysis methods, as we are interested in a *time-to-event* ("event" = severe COVID-19 or censoring)

Cox proportional hazards $h(t) = h_0(t)\exp(X\beta)$

with L1 regularization (***Lasso-Cox***) $\lambda\left|\left|\beta\right|\right|_1$

(λ selected using grid search with 5-fold cross validation, optimizing for discriminative ability as measured by concordance, or *C-index*)

Experimental Setup: Feature Sets

Lasso-Cox models learned from five different feature sets:

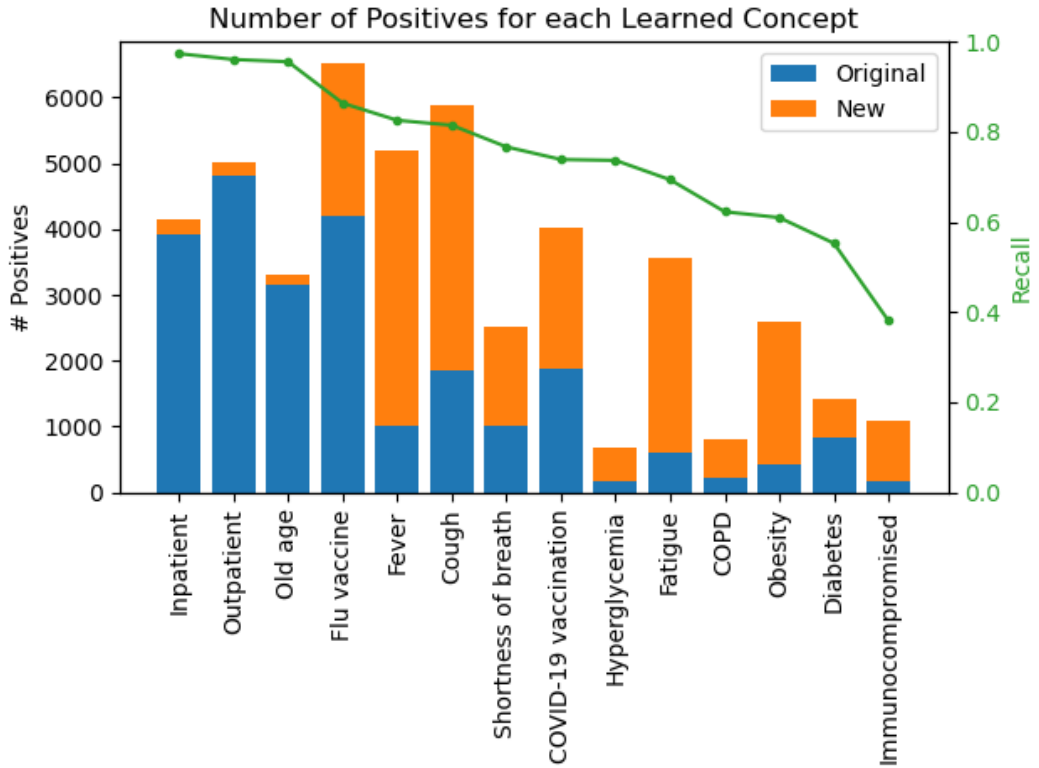
1. **Raw positive anchors:** without learning the corresponding clinical concepts
2. **Learned concepts (LC):** only the concepts from PU learning
3. **LC + Numeric:** learned concepts + numerical features
4. **LC + All features:** learned concepts + all of the original 139 features
5. **All features:** all 139 original features, no learned concepts

Experimental Setup: Data Splits

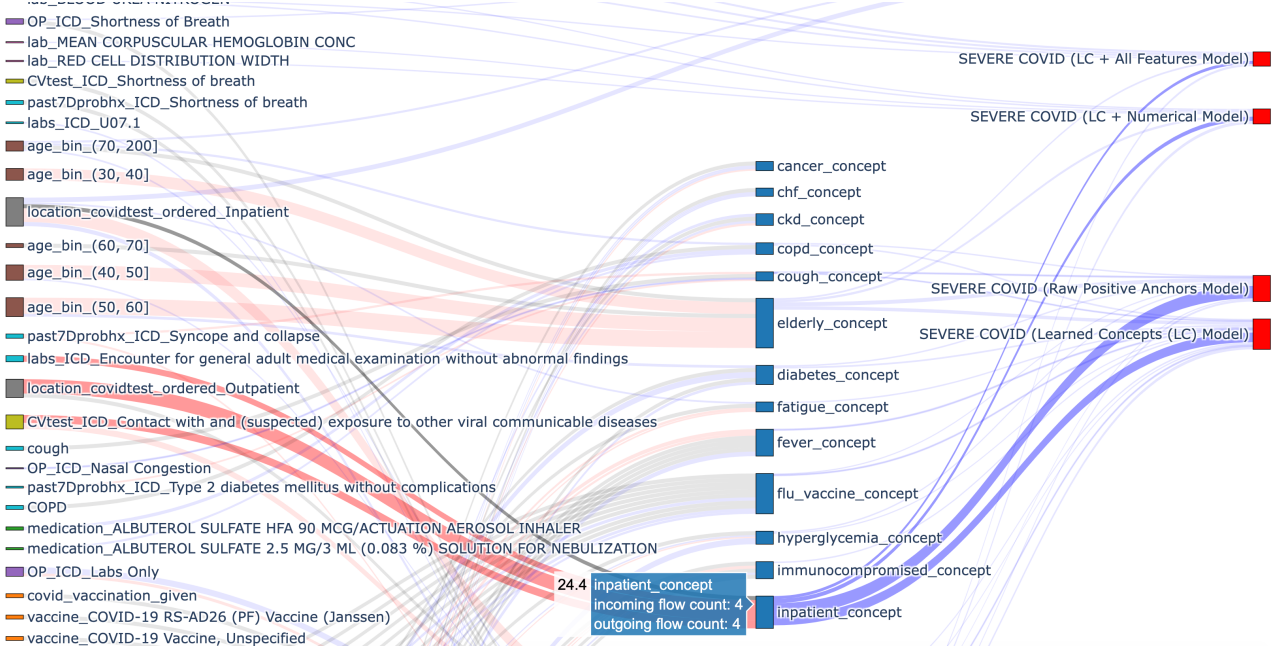
As data continues to be generated, hospitals may use new data to update their models over time.

- Performance over time setup:
  - Re-train models up to the end of each season (spring, summer, fall, winter)
  - Evaluate on subsequent seasons
  - 70-30 split to create train/test sets for each 3-month period
- Standard setup:
  - In addition, we train a model on the entire study time range
  - Train and test sets aggregate the respective 3-month datasets

Results: Learned Concepts

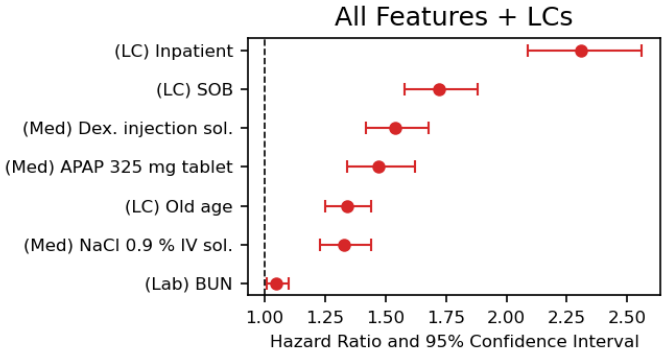


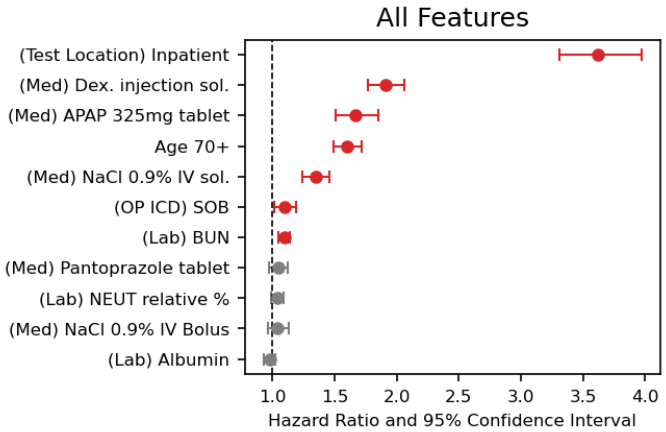
Interactive Sankey Diagram

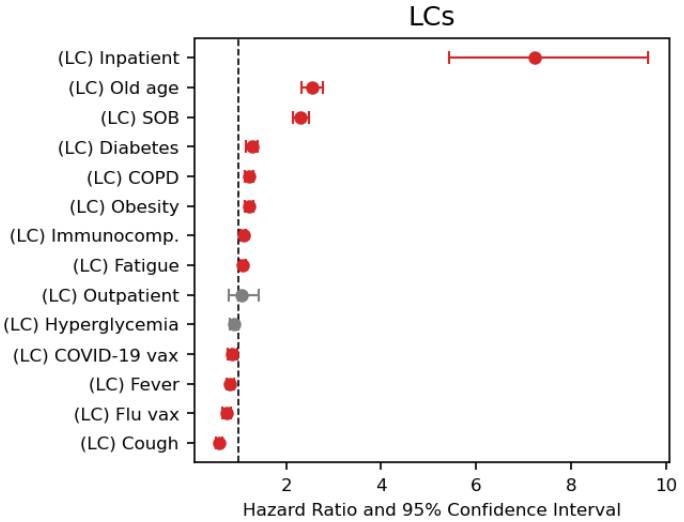


acmilab.org/severe_covid

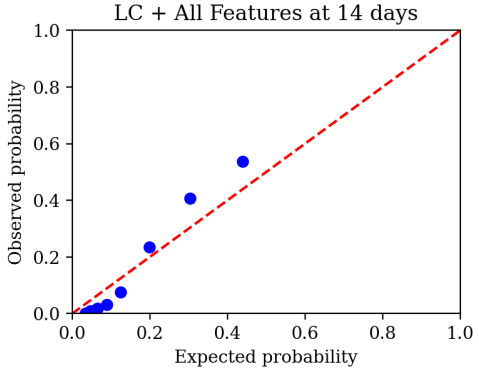
Results: Hazard Ratios

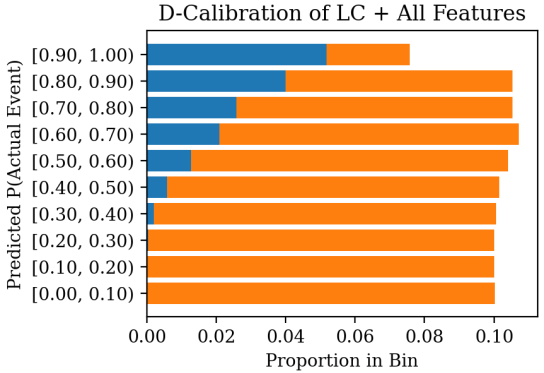


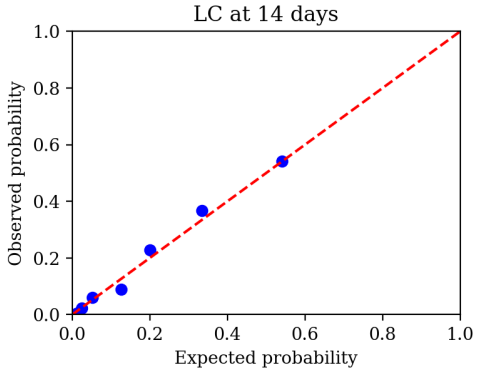


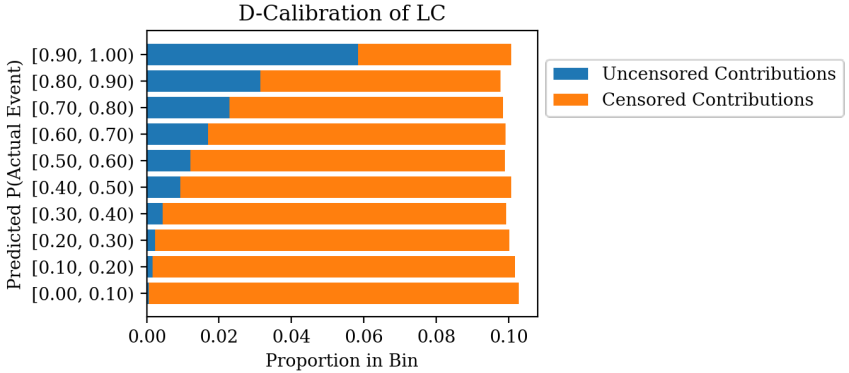


Calibration

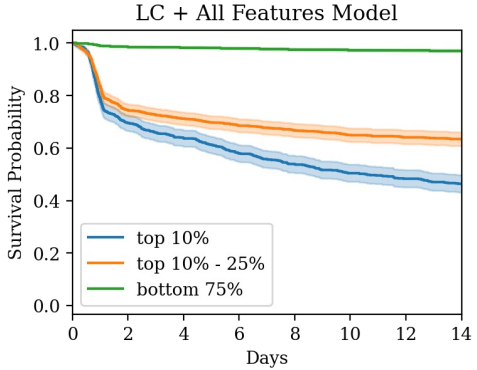




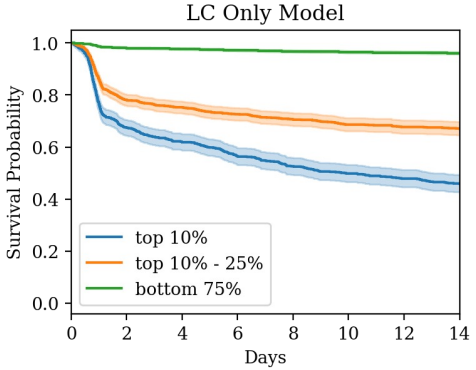




Kaplan Meier Survival Curves



top 10%								
At risk	938	627	558	496	449	408	380	358
Censored	0	28	45	56	68	81	92	98
Events	0	283	335	386	421	449	466	482
top 10% - 25%								
At risk	1406	967	874	804	755	715	698	673
Censored	0	86	137	174	202	223	230	247
Events	0	353	395	428	449	468	478	486
bottom 75%								
At risk	7030	6220	5921	5642	5458	5332	5205	5097
Censored	0	709	994	1250	1412	1526	1644	1742
Events	0	101	115	138	160	172	181	191

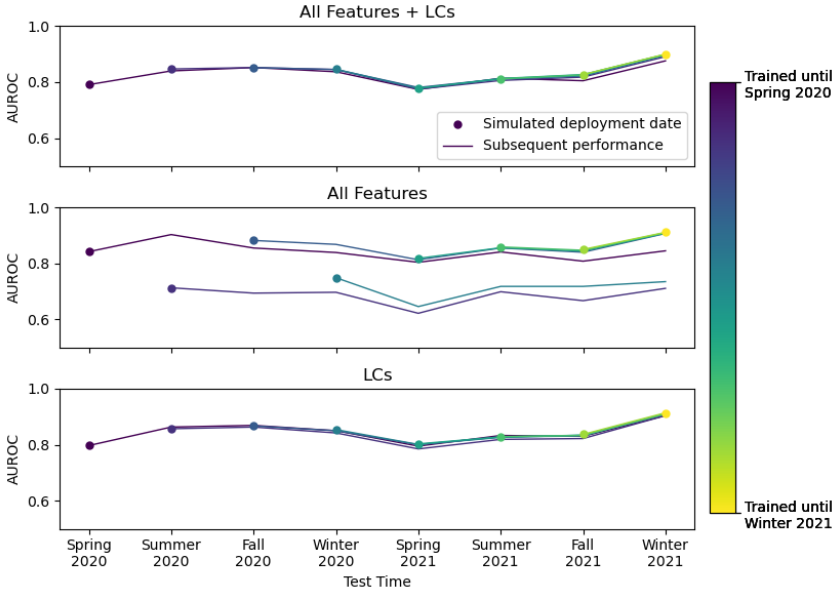


top 10%								
At risk	938	595	520	458	418	383	359	335
Censored	0	41	67	83	93	106	115	124
Events	0	302	351	397	427	449	464	479
top 10% - 25%								
At risk	1406	991	911	838	785	747	726	705
Censored	0	115	159	200	229	246	257	270
Events	0	300	336	368	392	413	423	431
bottom 75%								
At risk	7030	6228	5922	5646	5459	5325	5198	5088
Censored	0	667	950	1197	1360	1478	1594	1693
Events	0	135	158	187	211	227	238	249

Results: Aggregate Performance

Model	Aggregate Test C-index	Inpatient Test C-index	Outpatient Test C-index
Covichem	0.60 (0.58–0.62)	0.58 (0.57–0.60)	0.55 (0.51–0.58)
Galloway count	0.75 (0.73–0.76)	0.65 (0.63–0.66)	0.71 (0.68–0.75)
Galloway reweighted	0.81 (0.80–0.82)	0.70 (0.67–0.70)	0.76 (0.73–0.71)
Raw positive anchors	0.84 (0.84–0.85)	0.67 (0.65–0.68)	0.76 (0.71–0.80)
LC only	0.86 (0.85–0.87)	0.70 (0.69–0.71)	0.80 (0.76–0.83)
LC + numerical	0.86 (0.85–0.87)	0.70 (0.68–0.71)	0.81 (0.78–0.85)
LC + all features	**0.87 (0.87–0.88)**	**0.72 (0.70–0.73)**	**0.88 (0.86–0.90)**
All features (no LC)	**0.87 (0.87–0.88)**	**0.72 (0.70–0.73)**	**0.88 (0.86–0.90)**

Back-testing



Models trained in Spring 2020:

Feature Set	Spring 2020	Winter 2021
All Features + LCs	0.79	0.88
All Features	**0.84**	0.85
LCs	0.80	**0.90**

Summary

- Learned concepts anchored to clinical intuition
- Utilized learned concepts for downstream severe COVID-19 prediction
- Including all features boosts test performance when evaluated in aggregate
- But learned concepts resulted in more robust performance over time
- Visualization tool for examining precisely how the concepts are formed

Thank you!

Email me at:
hlzhou@cmu.edu