Factors Associated with Emergency Department Return Among COVID-19 Patients

PI: Jessica Galarraga, MD, MPH
Allan Fong, MS., Katharine T. Adams, MS., Erin Giovannetti, PhD,
Gongliang Zhang, PhD, MS., Hayley T. Rogovin

Medstar Health
UNC

Project Dates: 09/30/2021 – 09/30/2023
Federal Project Officer: Dr. Mario Terán, MD
This work was supported by the Agency for Healthcare Research and Quality.
Grant Award Number: R21 HS028563-01

STRUCTURED ABSTRACT

Purpose: Our objective was to develop machine learning (ML) algorithms to predict the risk for ED return and morbidity or mortality among returns. These algorithms can inform clinicians on whether admission to the hospital may be needed to prevent adverse outcomes among COVID-19 patients and identify patients who are safe for discharge to reserve hospital resources for those at greatest risk for failed outpatient management.

Scope: The focus of this grant was to (Aim 1) iteratively develop a concept map using mixed methods to identify predictive factors for COVID-19 ED returns. These factors were then used to inform (Aim 2) the development and evaluation of ML algorithms predictive of ED return risk for COVID-19 patients, and to (Aim 3) prospectively validate a the COVID-19 ED return model.

Methods: We took a mixed methods approach. Aim 1 entailed a thematic analysis of semi-structured interviews, with the interview approach informed by concepts identified in bi-variate logistic regression and natural language processing of EHR documentation. Aim 2 and Aim 3 involved data mining and the use of ML to develop multiple COVID-19 ED return predictive models trained on four different feature types was compared: clinical EHR data, word embeddings derived from applying an NLP BERT model to clinical documents, bag-of-words features derived from clinical documents, and a combination of all features. Health Information Exchange (HIE) data was incorporated into the predicted variable calculation and subsequent model performance was evaluated. SHAP value analysis was also applied.

Results: We generated a concept map, developed a ML model, and applied that model across sites with near real-time data. The final dataset consisted of 26,454 encounters representing 26,454 unique patients with a 9.9% rate of 9-day ED return. The highest ROCAUC for the primary data was 0.659 for a model trained on clinical EHR data features only, followed by all features (0.651), the bag-of-words model (0.611), and the BERT embeddings model (0.551). ROCAUC, F1 score, and precision-recall AUC improved in all feature types when using the expanded HIE target variable in training and testing. The HIE target variable model trained on clinical EHR data had the highest ROCAUC (0.671).

Keywords: machine learning, natural language processing, COVID-19, emergency department returns

SECTION I: PURPOSE

Objectives of the study

Our objective was to develop ML algorithms to predict the risk for ED return and morbidity or mortality among returns. These algorithms can inform clinicians on whether admission to the hospital may be needed to prevent adverse outcomes among COVID-19 patients and identify patients who are safe for discharge to reserve hospital resources for those at greatest risk for failed outpatient management.

The project included three aims:

- Aim 1: Iteratively develop a concept map using mixed methods, which serves as the ontology categorizing predictive factors for COVID-19 ED returns to inform ML model development.
    - *Subaim 1a:* Apply natural language processing (NLP) algorithms to EHR free-text notes of the COVID-19 ED population to develop the initial concept map. Employ unsupervised clustering techniques to identify features associated with COVID-19 ED returns, including clinical and social factors.
    - *Subaim 1b:* Conduct semi-structured interviews with care transition and clinical experts, perform a thematic analysis, and incorporate emerging themes into the iterative refinements of the concept map.

- Aim 2: Develop and evaluate ML algorithms predictive of ED return risk for COVID-19 patients.
    - Develop ML algorithms to predict the risk of ED return and, among returns, risk of hospitalization and mortality for COVID-19 patients. Use EHR data, including free-text notes with NLP, to train and test the models. Test model performance with a different EHR platform from a second health system and with the inclusion of data from the regional health information exchange to comprehensively detect ED returns.

- Aim 3: Prospectively validate a *C*OVID-19 *E*D *r*eturn *s*creening *t*ool (CERST) using real-time data.
    - Perform phase one of CERST implementation by prospectively testing and optimizing model performance with real-time data monitoring and analysis of EHR data.

SECTION II: SCOPE

Background

The novel coronavirus disease-2019 (COVID-19) pandemic placed unprecedented demands on emergency departments (EDs) to evaluate and treat large volumes of patients with suspected severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. Moreover, the array of COVID-19 clinical presentations is extensive, with highly variable symptoms and the course of disease ranging from uneventful recovery to multi-organ system failure and death. There is a need to develop predictive models that can support clinical decision making to determine appropriate and safe dispositions for COVID-19 patients in the ED setting, where resources are often constrained and patients present early on during their COVID-19 illness when their clinical course trajectory is most unpredictable.

The recent COVID-19 pandemic created an imperative to predict patient outcomes in the emergency department (ED) setting for which ML is a potential solution. ML and NLP can leverage the data-rich environment of the ED to provide timely guidance for safe dispositions of COVID-19 patients, which may improve quality in a highly resource-constrained clinical setting.

Context

The application of predictive machine learning (ML) algorithms can address this important healthcare challenge on the frontline pandemic response and improve the quality of care delivered to patients with COVID-19, a condition that has demonstrated to have continued clinical relevance post-pandemic. ML enables us to harness large amounts of dynamic and heterogeneous data generated from electronic health records (EHRs) to develop predictive tools that can provide timely guidance for emergency clinicians managing this disease. Using ML algorithms to predict the risk for ED return and morbidity or mortality among returns can inform clinicians on whether admission to the hospital may be needed to prevent adverse outcomes among COVID-19 patients and identify patients who are safe for discharge to reserve hospital resources for those at greatest risk for failed outpatient management. This project develops the foundation for a screening tool using EHR data to support safe COVID-19 discharges in the ED setting.

Settings

The project included data for model training and testing from two large regional health systems, MedStar Health (MSH) and University of North Carolina Health (UNC), to support the generalizability of project findings. There are a total of 19 acute care hospitals represented by these two health systems, which capture a broad patient population across Maryland, the District of Columbia (DC), and North Carolina. They also represent a diversity of sites, including urban, suburban, rural, academic, and community sites, and the two health systems use different EHRs, Cerner at MSH and Epic at UNC, further enhancing the project’s generalizability.

Participants

The study population for the modeling aims included all index ED visits by adult COVID-19 patients, ≥ 18 years, who were discharged home over an 18-month period, April 1, 2020 to September 30, 2021. COVID-19 positive patients were defined based on clinical diagnosis and/or a positive COVID-19 laboratory test. We focused on adult patients since the clinical features, course trajectory, and risk factors for COVID-19 illness differ for the pediatric population and pediatric patients are not consistently present among the sites in the health systems providing source data. We excluded index ED encounters that resulted in the patient leaving against medical advice or eloping since the disposition decision is not relevant to those encounters and the algorithm development focuses on ED patients who are discharged home to determine a risk profile that is relevant to clinical workflows for disposition decision making when the screening tool is operationalized.

SECTION III: METHODS

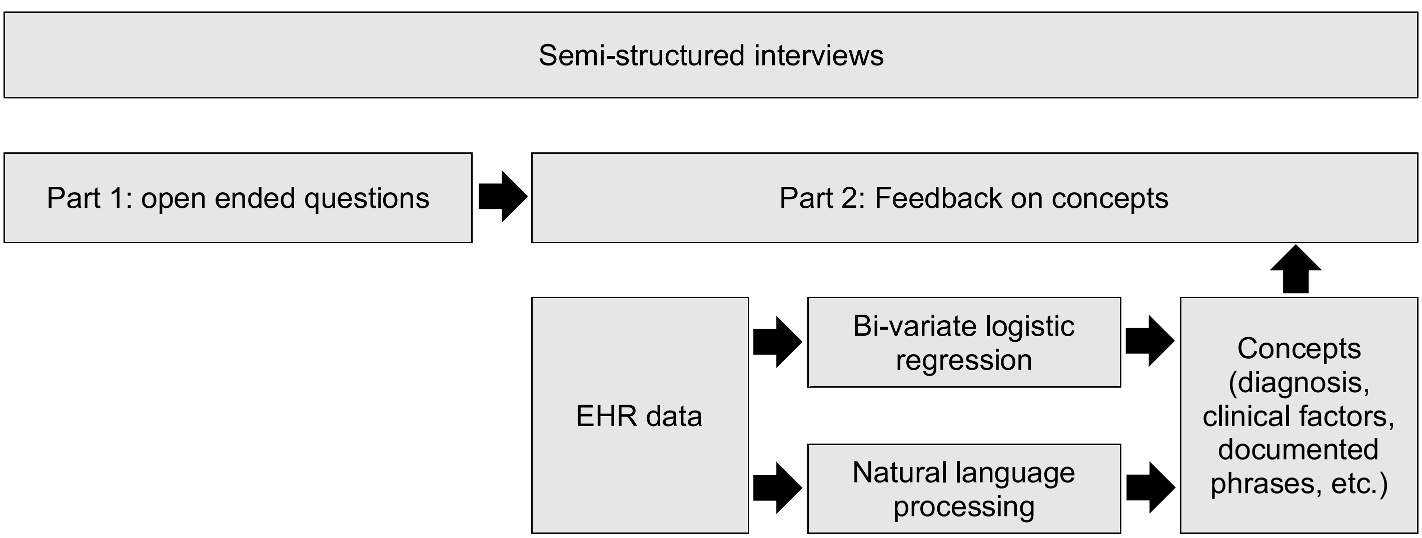
We first describe the dataset that was created and used to address Aims 2 and 3. We then describe the various methods used to support each aim, other than the ED dataset.

Emergency department dataset. Our cohort consisted of COVID-19 patients 18+ years who utilized the ED across 9 MedStar sites between 4/1/2020 and 9/30/2022 and the target indicator was an ED return within 9 days of the initial visit. Over 1 million EHR data points were retrieved including labs, vitals, medications, comorbidities, diagnoses, demographics, imaging, ED notes, and radiology reports. Our final dataset consisted of 21,780 encounters with a 9-day return rate of 10.2%. Predictive models trained using the XGBoost algorithm and built on different feature sets were created and compared, including structured EHR data, structured EHR data mapped to concepts from qualitative interviews with subject matter experts (SMEs), and NLP features constructed from clinical notes. Feature importance was calculated to glean additional insight into relevant features driving model decisions.

Aim 1a. We performed a retrospective evaluation of ED encounters among adult COVID-19 patients who were discharged home using electronic health record structured and unstructured data across nine MedStar Health EDs, from April 1, 2020 to March 31, 2022. We used bivariate logistic regression to identify factors associated with ED return, including patient and encounter characteristics, vital signs, and laboratory values. We also used natural language processing (NLP) to perform unsupervised clustering of clinical free-text notes, to identify features associated with ED return. The ED return measure was tested at both 72-hours and 9-days.

Aim 1b. To develop the concept map, semi-structured interviews were performed with clinicians and care coordination staff. Interviews included open-ended questions eliciting observations of COVID-19 patients and return encounters, followed by structured feedback in response to the list of concepts identified in the initial quantitative analysis. Figure 1 provides a summary of this mixed methods approach. We conducted interviews until saturation was reached in both groups of participants (clinicians and care coordination staff), defined as no new emergent themes. We performed thematic analysis of open-ended responses combined with descriptive statistics of structured responses to develop a concept map.

Figure 1: Mixed Methods Approach to Concept Map Development



Aim 2. Machine learning models were trained using four different feature sets: 1) structured EHR data, 2) bag-of-words bigram binary tokens and a logic-based algorithm to identify positive radiology results, 3) clinical BERT embeddings concatenated from four different text types, and 4) all features.

Structured EHR Data
Patient and visit-level data were collected from the electronic health record (EHR). Patient information used as model features included the patient's age at the time of the visit, smoking status, body-mass index (BMI), sex, insurance, race and ethnicity, the patient's national ADI percentile based on current zip code, a binary variable indicating whether the patient retained a primary care physician, and a categorical representation of the patient's number of ED visits in the past year.

Index visit information included the patient's Emergency Severity Index (ESI) at the time of triage, the mode of arrival to the ED, index visit diagnosis data, medications ordered, administered, or recorded as a home medication at the visit, a binary variable indicating whether the patient received an EKG at the visit, and imaging data. Binary variables indicating the administration of COVID-19 oral antivirals and monoclonal antibodies at the index visit were included. Encounters missing one or more of these variables were removed from the final data, resulting in 629 excluded encounters.

Data from thirty-four unique lab tests and seven vital signs were retrieved from the EHR. Maximum, minimum, average, first, last, and the maximum value within an hour of discharge were calculated for each lab and vital sign. The total number of lab and vital results outside of the normal range per unique lab test was also calculated. The lab normal ranges are provided by the EHR and actively used in clinical decision support and front-end clinical representations in the healthcare system. Given the relative infrequency but clinical significance of a patient's ambulatory SPO2, a binary variable indicating whether an SPO2 result was recorded during the index encounter was calculated. Encounters with missing one of the six main vital signs, excluding ambulatory SPO2, were removed from the final data, resulting in 1,058 excluded encounters.

Diagnoses and medications associated with a patient's prior visits within the past year were retrieved. Twenty-one patient comorbidities were defined using diagnosis recorded within the past year and grouped by subject matter experts (SMEs). Finally, binary variables indicating historic administration of COVID-19 oral antivirals or monoclonal antibodies within the past year were created. Encounters without at least once diagnosis during the index visit were excluded from the final data, resulting in 38 excluded encounters.

Unstructured EHR Data
Text features were extracted from two different EHR sources: ED clinical notes and radiology reports. Each ED clinical note was split into its structured subsections using rule-based logic, with the History of Present Illness (HPI), Assessment/Plan (AP), and ED Course/Critical Care (EDCCC) sections retained separately for further model experimentation. Only the findings/impressions and results sections of the radiology notes were retained and kept as a single text. Each text was converted to lowercase and then cleaned by removing punctuation, white space, and line breaks.

A pretrained BERT tokenizer and model trained on clinical discharge summaries from the MIMIC database was applied to each of the four text types. As the clinical BERT model has a maximum sequence length of 128, the texts were truncated and tokenized using the pretrained tokenizer and an attention mask was applied where relevant. The embeddings of the first classification token [CLS] for the sequence of hidden states in the final layer of the BERT output was used as the final representation of the text, resulting in a 768-feature dataset per clinical text.

We preprocessed the ED and radiology notes by removing stopwords and extracting relevant content by note section. We extracted bigrams and calculated pointwise mutual information (PMI) for each bigram. PMI is an association metric that compares the probability of two things occurring together with the probability of them occurring independently. Bigrams were selected based on top quantile of PMI scores for ED and radiology tokens separately. Bigrams relevant to the radiology and full ED note text were extracted as individual features.

A positive indicator of abnormality in the radiology text was derived using a logic-based extraction algorithm derived from Subject Matter Expert (SME) input and used as a feature.

Primary target variable: ED return
The four models were trained on the 9-day ED return as the target variable, including an HIE expanded 9-day ED return indicator. Preliminary analysis examined the 72-hour versus 9-day ED return metrics and demonstrated better performance with 9-day ED returns. This is consistent with prior research that has demonstrated that 9 days is the time metric for ED returns that most accurately captures the majority of patients returning to the ED as a result of post-discharge complications or outpatient treatment failure.
Model performance was determined by standard machine-learning metrics: ROCAUC, F1 score, precision, and recall. Shapley Additive exPlanations (SHAP) analysis was applied to the final model to help visualize feature contribution.

Secondary target variables: ED return hospitalization and ED return mortality
Models were trained on two secondary target variables 1) hospitalization within 9 days of discharge from the ED index visit, and 2) mortality within 9 days of discharge from the ED index visit. The incidence of positive cases in our patient cohort for these two secondary target variables was limited, 3.1% of index encounters resulted in a hospitalization within 9 days of discharge and 0.2% of index encounters resulted in patient mortality within 9 days of discharge. To address the imbalance, models were trained using a weighted positive class parameter that reflected the ratio of positive cases to negative cases. Models trained only on structured EHR data were the best performing and those trained on BERT embeddings were the lowest performers for both secondary indicators.

Multisystem testing
To evaluate generalizability, Extreme Gradient Boosting (XGBoost) models were trained and tested on a total of 42,056 encounters from MSH (HS1) and UNC (HS2) health systems: HS1 (26,454 encounters) and HS2 (16,602 encounters). 3537 features were used in the model training. 2920 were diagnoses recorded at the index visit, 489 were medications, 67 were laboratory value-based features, 22 were patient-level features such as demographics, insurance, BMI, 18 were patient comorbidities, 18 were vital sign-based features, and 3 were historical utilization patterns. Five cross-fold validation was used for training on 80% of the data. We tested the models on the held-out 20%, which was evaluated for area under the receiver operating characteristic curve (ROCAUC), recall, precision, F1 score, and area under the precision recall curve (PRAUC). We compared the results by health system and evaluated feature importance using Shapley Additive exPlanations (SHAP) and Gini values.

Aim 3. We applied the model to real-time data on a weekly level and review per week. We prospectively validated a machine learning model that would support a *C*OVID-19 *E*D *r*eturn *s*creening *t*ool (CERST) using real-time data. This supported phase one of CERST implementation by prospectively testing and optimizing model performance with real-time data monitoring and analysis of EHR data.

Section IV: Results

Aim 1a. The study sample included 24,940 COVID-19 ED encounters, with a 72-hour ED return rate of 6% and 9-day ED return rate of 12%. Factors identified from EHR structured data associated with increased odds of 9-day ED return include older age, male sex, higher body mass index, having liver disease/cirrhosis, having a longer ED length of stay, and prior ED utilization. (Table 1)

Table 1. Bivariate Logistic Regression of Patient and Encounter Characteristics with Odds of 9-day ED return among COVID-19 encounters, April 1, 2020 – March 31, 2022

Study Variable	Odds Ratio	P-value
Patient Characteristics		
Age (years)	1.01	<0.001
Biological Sex (Ref = Male)		
Female*	0.90	0.02
Body Mass Index (Ref = 18.5 - 25)		
< 18.5	1.20	0.29
>= 25, < 30*	1.19	0.00
>= 30, < 35*	1.29	<0.001
>= 35, < 40*	1.24	0.01
>= 40*	1.33	0.00
Race and Ethnicity (Ref = Non-Hispanic White)		
Hispanic/Latino	1.14	0.26
Non-Hispanic Black	1.08	0.15
Other race*	1.19	0.02
Primary Insurance (Ref = Private)		
Medicaid Only	1.05	0.33
Medicare/Medicaid Dual Eligible	0.91	0.34
Medicare Only	1.12	0.19
Other	1.16	0.17
Self-pay/uninsured	1.00	0.99
Smoking Status (Ref = Never Smoker)		
Current smoker	0.89	0.08
Former smoker	1.12	0.12
Unknown*	0.66	<0.001
Comorbidities (Ref = Comorbidity Absent)		
Asthma	0.95	0.41
Bone Marrow Transplant	1.40	0.09
Coronary Artery Disease	0.92	0.38
Chronic Obstructive Pulmonary Disease	1.21	0.08
Cardiomyopathy	0.82	0.28
Cerebrovascular Disease	1.19	0.13
Chronic Kidney Disease	1.17	0.19
Congestive Heart Failure	0.93	0.61
Dementia	0.92	0.63
Diabetes, Type 1 or 2	1.08	0.24
HIV	0.99	0.95
Hypertension	1.05	0.43
Liver Disease/Cirrhosis*	1.46	<0.001
Obesity	0.96	0.42
Pulmonary Fibrosis	1.12	0.41
Sickle Cell Disease	1.06	0.81
Encounter Characteristics		
LOS (minutes)*	1.00	0.04
# ED Visits in past year*	1.13	<0.001

Ref = Reference group for categorical variables. ED = Emergency Department. All vital sign and laboratory variables used the average value for the index encounter. *p-value ≤ .05

Vital sign and laboratory results were notable for the following factors associated with ED return: higher heart rate, higher temperature, higher creatinine, lower lymphocytes, lower platelets, and lower pCO2 venous. (Table 2)

Table 2. Bivariate Logistic Regression Analysis of Clinical Features with Odds of 9-day ED return among COVID-19 encounters, April 1, 2020 – March 31, 2022

Study Variable	Odds Ratio	P-value
Vital Signs		
Systolic Blood Pressure	1.00	0.63
Diastolic Blood Pressure*	0.99	0.01
Respiratory Rate	1.02	0.09
SpO2	0.99	0.24
Ambulatory SpO2	1.01	0.20
Temperature*	1.14	<0.001
Heart Rate*	1.01	<0.001
Laboratory Values		
Anion Gap	0.99	0.43
ALT	1.00	0.71
AST	1.00	0.44
Amylase	0.98	0.13
BUN	1.00	0.96
CO2	0.98	0.10
CRP	1.01	0.34
CRP_High Sensitivity	1.00	0.71
Creatinine Kinase*	1.00	0.04
Creatinine*	1.05	0.05
D-dimer	0.74	0.88
Ferritin	1.00	0.76
Fibrinogen	0.99	0.24
Glucose (Random)	1.00	0.64
Hgb	0.98	0.31
HbA1C	0.97	0.87
INR	0.92	0.61
LDH	1.00	0.09
Lactic Acid	0.97	0.84
Lipase	1.00	0.33
Lymphocyte %*	0.99	0.03
Lymphocyte Absolute	1.00	0.97
NTproBNP	1.00	0.66
Platelet*	0.998	<0.001
Procalcitonin	0.01	1.00
Troponin I	0.01	0.15
WBC	0.97	0.12
pCO2_Venous*	0.93	0.004
pO2_Art	0.99	0.59
pH_Arterial	1.00	0.49
pH_Venous	2.09	0.82

Ref = Reference group for categorical variables. ED = Emergency Department.
All vital sign and laboratory variables used the average value for the index encounter. *p-value ≤ .05

Table 3: Summarizes NLP features associated with ED return. The count of the feature, individual word counts, and pairwise mutual information (PMI) scores are provided.

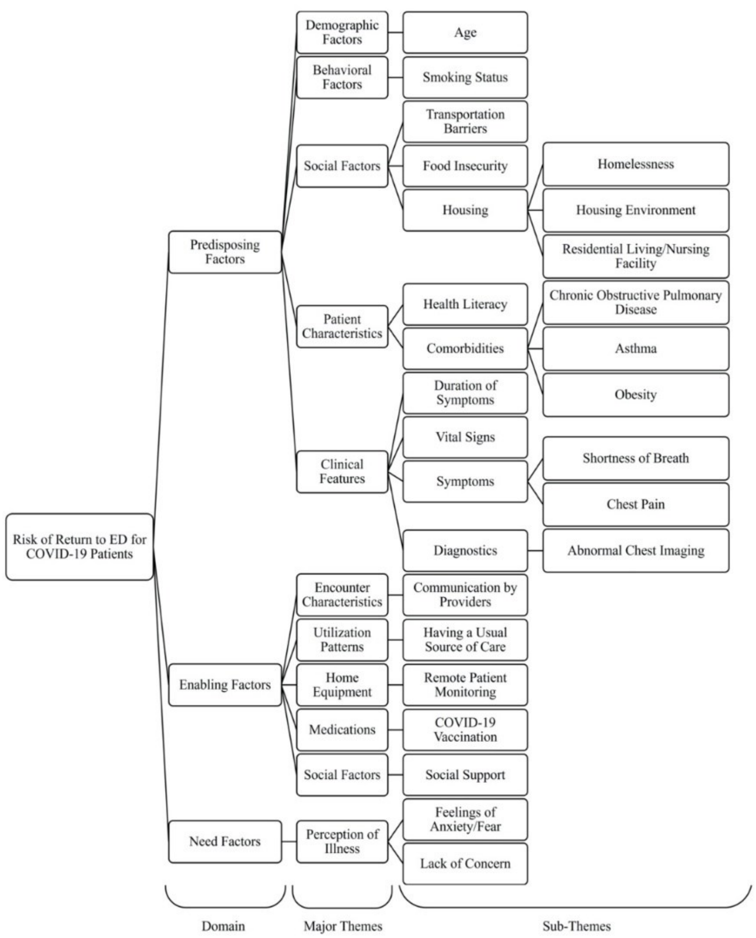
Feature	Count	word 1	word 2	w1_count	w2_count	pmi
bipolar disorder	385	bipolar	disorder	639	1254	8465.08
inhalation aerosol	855	inhalation	aerosol	2181	861	8021.88
alcohol withdrawal	339	alcohol	withdrawal	5916	521	1937.77
alcohol abuse	294	alcohol	abuse	5916	2088	419.33
feeling lightheaded	262	feeling	lightheaded	8169	1609	351.19
viral pneumonia	383	viral	pneumonia	5668	14228	83.67
covid pneumonia	2236	covid	pneumonia	105864	14228	26.15
sob days	245	sob	days	15084	36513	7.84
with fever	1507	with	fever	177079	34657	4.33
sob with	321	sob	with	15084	177079	2.12

Additionally, the NLP analysis found the following concepts to be associated with ED return: symptoms of severe cough, dehydration, respiratory difficulties, or wheezing, the patient having uncontrolled diabetes, clinical documentation of shared decision making, and clinical documentation of anticipatory guidance. (Table 3)

Aim 1b. 17 interviews were conducted with ED clinicians (n=11) and care coordination staff (n=6) representing seven sites. The sites represented by interview participants included four inner city, two suburban, and one rural ED, with an annual ED volume ranging from 8,000 to 90,000 encounters. Two sites were teaching hospitals and five were non-teaching hospitals. The most frequently reported perceived risk factors for COVID-19 ED return included social factors (N=10), patient perceptions of their illness (N=7), and comorbidities (N=7). Social factors included housing environment (e.g., crowdedness, homelessness), transportation barriers, and food insecurity, and the most reported comorbidities were COPD and asthma. Factors rated by clinicians as most strongly associated with ED return were comorbidities, history of high ED use, low oxygen saturation, and symptoms of chest pain. Care coordination staff additionally rated immunosuppression, not having health insurance, and unmet social needs as having a very strong association. The two enabling factors identified by both clinicians and care coordination staff to be strongly associated with decreased risk of ED return in COVID-19 patients were remote patient monitoring and, along similar lines, home pulse oximeter use. The concept map for COVID-19 ED returns produced from the mixed methods analysis, using the Anderson-Behavioral model, is summarized in Figure 2.

The primary perceived factors associated with an increased risk for ED return resulting in hospitalization or mortality were comorbidities, including COPD, asthma, and CHF, social factors, including social support and housing environment, medications, including COVID-19 vaccination and immunosuppressant use, and age.

Figure 2. Emerging themes associated with risk of Emergency Department Return for COVID-19 Patients



Aim 2: The final dataset for ML model training and testing consisted of 26,454 encounters representing 26,454 unique patients with a 9.9% rate of 9-day ED return. The highest ROCAUC for the primary data was 0.659 for a model trained on clinical EHR data features only, followed by all features (0.651), the bag-of-words model (0.611), and the BERT embeddings model (0.551). ROCAUC, F1 score, and precision-recall AUC improved in all feature types when using the expanded HIE target variable in training and testing. (Table 4)

Table 4. 9-day ED return model metrics for training and test data for all four predictive models

				ROCAUC				F1				Precision				Recall			
Feature Type	% Return	Number of Encounters	Number of Features	Train	Test			Train	Test			Train	Test			Train	Test		
					Mean	95% CI	STD		Mean	95% CI	STD		Mean	95% CI	STD		Mean	95% CI	STD
Structured EHR Data	9.9%	26454	10464	0.717	0.664	[0.656 - 0.671]	0.006	0.279	0.25	[0.24 - 0.26]	0.007	0.179	0.16	[0.154 - 0.166]	0.005	0.631	0.567	[0.54 - 0.594]	0.02
NLP: BERT Embeddings	10.3%	23863	3072	0.924	0.556	[0.534 - 0.578]	0.016	0.507	0.186	[0.166 - 0.205]	0.014	0.359	0.132	[0.118 - 0.147]	0.011	0.858	0.311	[0.275 - 0.346]	0.026
NLP: Bag-of-Words + Logic	9.9%	26454	620	0.673	0.618	[0.607 - 0.635]	0.012	0.249	0.221	[0.208 - 0.235]	0.01	0.157	0.14	[0.131 - 0.148]	0.006	0.598	0.532	[0.498 - 0.566]	0.025
All Features	9.9%	26454	14156	0.709	0.656	[0.649 - 0.664]	0.005	0.273	0.241	[0.237 - 0.244]	0.002	0.174	0.154	[0.151 - 0.157]	0.002	0.626	0.552	[0.546 - 0.557]	0.004

The HIE target variable model trained on clinical EHR data had the highest ROCAUC (0.671). (Table 5)

Table 5. Model metric comparison between original and HIE-augmented target variable

				Original Target Variable ROCAUC				HIE Target Variable ROCAUC			
Feature Type	% Return	Number of Encounters	Number of Features	Train	Test			Train	Test		
					Mean	95% CI	STD		Mean	95% CI	STD
Structured EHR Data	9.9% 11.9%	26454	10464	0.717	0.664	[0.656 - 0.671]	0.006	0.717	**0.675**	[0.662 - 0.687]	0.009
NLP: BERT Embeddings	10.3% 12.5%	23863	3072	0.924	0.556	[0.534 - 0.578]	0.016	0.904	**0.584**	[0.564 - 0.603]	0.014
NLP: Bag-of-Words + Logic	9.9% 11.9%	26454	620	0.673	0.618	[0.607 - 0.635]	0.012	0.675	0.629	[0.62 - 0.637]	0.006
All Features	9.9% 11.9%	26454	14156	0.709	0.656	[0.649 - 0.664]	0.005	0.706	**0.669**	[0.662 - 0.677]	0.005
				Original Target Variable F1				HIE Target Variable F1			
Feature Type	% Return	Number of Encounters	Number of Features	Train	Test			Train	Test		
					Mean	95% CI	STD		Mean	95% CI	STD
Structured EHR Data	9.9% 11.9%	26454	10464	0.279	0.25	[0.24 - 0.26]	0.007	0.318	**0.289**	[0.281 - 0.297]	0.006
NLP: BERT Embeddings	10.3% 12.5%	23863	3072	0.507	0.186	[0.166 - 0.205]	0.014	0.53	**0.23**	[0.208 - 0.252]	0.016
NLP: Bag-of-Words + Logic	9.9% 11.9%	26454	620	0.249	0.221	[0.208 - 0.235]	0.01	0.285	**0.264**	[0.258 - 0.27]	0.004
All Features	9.9% 11.9%	26454	14156	0.273	0.241	[0.237 - 0.244]	0.002	0.31	**0.291**	[0.279 - 0.304]	0.009
				Original Target Variable Precision				HIE Target Variable Precision			

Feature Type	% Return	Number of Encounters	Number of Features	Train	Test			Train	Test		
					Mean	95% CI	STD		Mean	95% CI	STD
Structured EHR Data	9.9% 11.9%	26454	10464	0.179	0.16	[0.154 - 0.166]	0.005	0.212	**0.193**	[0.187 - 0.199]	0.004
NLP: BERT Embeddings	10.3% 12.5%	23863	3072	0.359	0.132	[0.118 - 0.147]	0.011	0.39	**0.17**	[0.154 - 0.187]	0.012
NLP: Bag-of-Words + Logic	9.9% 11.9%	26454	620	0.157	0.14	[0.131 - 0.148]	0.006	0.186	**0.172**	[0.168 - 0.177]	0.003
All Features	9.9% 11.9%	26454	14156	0.174	0.154	[0.151 - 0.157]	0.002	0.205	**0.193**	[0.184 - 0.202]	0.006
				Original Target Variable Recall				**HIE Target Variable Recall**			
Feature Type	% Return	Number of Encounters	Number of Features	Train	Test			Train	Test		
					Mean	95% CI	STD		Mean	95% CI	STD
Structured EHR Data	9.9% 11.9%	26454	10464	0.631	0.567	[0.54 - 0.594]	0.02	0.635	0.577	[0.562 - 0.593]	0.011
NLP: BERT Embeddings	10.3% 12.5%	23863	3072	0.858	0.311	[0.275 - 0.346]	0.026	0.827	**0.354**	[0.32 - 0.389]	0.025
NLP: Bag-of-Words + Logic	9.9% 11.9%	26454	620	0.598	0.532	[0.498 - 0.566]	0.025	0.608	0.565	[0.542 - 0.588]	0.017
All Features	9.9% 11.9%	26454	14156	0.626	0.552	[0.546 - 0.557]	0.004	0.634	**0.597**	[0.577 - 0.616]	0.014

The results of the SHAP value analysis for the structured EHR features (Figure 3) indicate that patients who have used the ED less than 3 times in the prior year are less likely to return to the ED. Heparin administered at a visit within the past year and glucocorticoids administered during the index visit indicate a higher likelihood of return. Higher values of a patient's age at the time of the index visit, a higher last recorded temperature, and lower values of SpO2 drove model prediction towards an ED return, while lower values of the last recorded heart rate had a stronger contribution towards a patient not returning. Finally, receiving monoclonal antibodies or oral antivirals to treat COVID-19 at the index visit was found by the model to contribute to a final prediction of not returning to the ED.

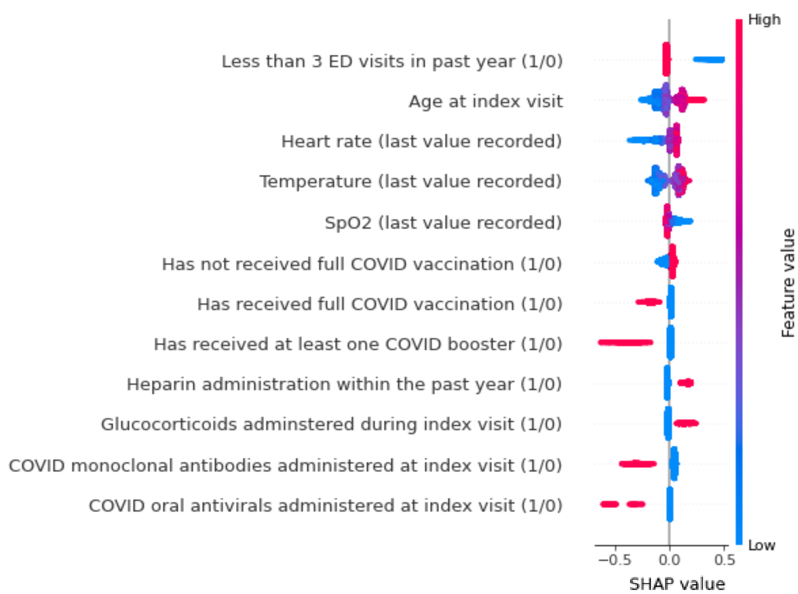


Figure 3. SHAP Values for Structured EHR Features

The SHAP results of the bag-of-words text-based model describe an interesting set of bigrams that impacted overall prediction. The phrase “medications inpatient” had the strongest contribution to model prediction, with the presence of the phrase indicating a likelihood to return. An indication of an abnormal radiology result derived from the radiology note strongly indicates the patient is more likely to return to the ED. Further, behavioral related phrases, such as “tobacco user”, “exercise frequency”, and “substance abuse” drove model prediction in an expected direction (“tobacco user” and “substance abuse” more likely to return, “exercise frequency” less likely).

9-Day ED Return Hospitalization and Mortality

The structured only and all feature models had similar performance for hospitalization prediction. (Table 7) The structured only model had the best ROCAUC performance for 9-Day mortality. (Table 9) Models trained only on structured EHR data were the best performing and those trained on BERT embeddings were the lowest performers for both secondary indicators. Full metrics are reported per indicator in the tables below. Additionally, descriptive tables per feature indicator are included that detail feature distribution differences for our hospitalization/no hospitalization cohort and mortality/no mortality cohort. In the case of categorical variables, the most frequent overall category is reported (i.e., female patients), continuous variables report means (patient age at the index encounter, number of ED visits in the past year, lab and vital signs). Feature categories with many binary variables (co-morbidities, documented medications) report the five most frequently documented features.

For patients that experienced a hospitalization within 9 days of the index ED encounter discharge, patients were older (mean 55.4 years vs. 44.3 years), were less likely to have a complete COVID vaccination (75.1% vs. 67.5%), more likely to have an EKG performed at the index encounter (64.7% vs. 44.4%), less likely to have arrived at the ED via walk-in (72.1% vs. 84.4%), and reported the highest difference in the average number of lab test results outside of the normal range for white blood cell, CO2, and creatinine lab tests compared to those that did not experience a hospitalization within 9 days of ED discharge. Additionally, hospitalized patients had a greater history of hypertension (44.4% vs. 26.8%) and Type I or II Diabetes (20.9% v. 11.7%) than non-hospitalized patients. (Table 6)

Table 6. Descriptives of Feature Indicators for 9-Day ED return hospitalization

Feature Category	**Feature(s)**	**No Hospitalization**	**Hospitalization**
		N = 25,636	**N = 818**
Demographic Factors	Age (mean)	44.3	55.4
	BMI: 25=< x <30	27.9%	30.2%
	Biological Sex: Female	59.1%	54.9%
	Insurance Status: Medicaid only	37.1%	24.8%
	Race/Ethnicity: Non-Hispanic Black	60.0%	54.6%
Utilization Patterns	Number of ED visits in the past year (mean)	0.71	0.91
Behavioral Factors	Smoking Status: Never Smoker	72.3%	72.4%
Documented Medications	Covid Vaccination Status: Not vaccinated or incomplete vaccination course	67.5%	75.1%
	Top 5 Medications Documented at Index Encounter	Misc. analgesics: 28% Nonsteroidal anti-inflammatory agents: 19.3% 5HT3 receptor antagonists: 13.2% Misc. antivirals: 12.5% Adrenergic bronchodilators: 0.9%	Misc. analgesics: 43.9% Nonsteroidal anti-inflammatory agents: 20.0% 5HT3 receptor antagonists: 19.9% Glucocorticoids: 13.6% Adrenergic bronchodilators: 12.6%
Diagnostic	EKG performed at index encounter	44.4%	64.7%
Encounter Acuity	ESI: 3V	31.4%	22.1%
	ESI (mean)	3.2	2.9

Feature Category	Feature(s)	No Hospitalization	Hospitalization
	Mode of Arrival: Walk-In	84.4%	72.1%
Comorbidities	Top 5 Comorbidities	Obesity: 31.2% Hypertension: 26.8% Asthma: 12.9% Diabetes Type I or II: 11.7% Coronary Artery Disease: 5.4%	Hypertension: 44.4% Obesity: 33.1% Diabetes Type I or II: 20.9% Coronary Artery Disease: 13.0% Chronic Kidney Disease: 12.3%
Vitals (last documented before discharge)	Diastolic Blood Pressure (mean)	77.8	76.0
	Systolic Blood Pressure (mean)	131.1	134.1
	Heart Rate (mean)	82.7	85.1
	Respiratory Rate (mean)	18.0	18.6
	Temperature (mean)	37.0	37.1
	SpO2 (mean)	97.8	96.9
	Ambulatory SpO2 (indicator, % documented during index visit)	21.7%	16.5%
Lab Tests – number of results outside normal range*	Hgb (mean)	0.14	0.22
	CO2 (mean)	0.22	0.36
	BUN (mean)	0.07	0.19
	Platelet (mean)	0.14	0.22
	White Blood Cell (mean)	0.15	0.31
	Creatinine (mean)	0.10	0.24
	AST (mean)	0.09	0.20
	Glucose Level Random (mean)	0.09	0.16
	ALT (mean)	0.17	0.26
	Neutrophil, Percentage (mean)	0.07	0.12
	Lymphocyte, Absolute (mean)	0.18	0.24

*Only lab values with the delta between the mean number of results outside of the normal range for hospitalized and non-hospitalized patients within 9 days of the index visit are reported.

Table 7: 9-day ED return hospitalization model results

				ROCAUC				F1			
Feature Type	% Return	Number of Encounters	Number of Features	Train	Test			Train	Test		
					Mean	95% CI	STD		Mean	95% CI	STD
Structured EHR Data	3.1%	26454	10464	0.863	0.807	[0.768 - 0.847]	0.029	0.17	0.151	[0.139 - 0.163]	0.009
NLP: BERT Embeddings	3.3%	23863	3072	0.937	0.69	[0.66 - 0.72]	0.022	0.251	0.12	[0.108 - 0.133]	0.009
NLP: Bag-of-Words + Logic	3.1%	26454	620	0.805	0.748	[0.728 - 0.769]	0.015	0.146	0.129	[0.118 - 0.141]	0.008
All Features	3.1%	26454	14156	0.931	0.8	[0.774 - 0.826]	0.019	0.237	0.161	[0.152 - 0.171]	0.007

For patients that experienced in-hospital mortality within 9 days of the index ED encounter discharge, patients were older (mean 64.4 years vs. 44.7 years), were less likely to have a complete COVID vaccination (76.7% vs. 67.7%), more likely to have an EKG performed at the index encounter (69.8% vs. 45.0%), less likely to have arrived at the ED via walk-in (65.1% vs. 84.1%), and reported the highest difference in the average number of lab test results outside of the normal range for random glucose levels, BUN, and creatinine range compared to those that did not experience in-hospital mortality within 9 days of ED discharge. (Table 8)

Table 8. Descriptives of Feature Indicators for 9-Day ED return mortality

Feature Category	Feature(s)	No Mortality	Mortality
		N = 26,411	N = 43
	Age (mean)	44.7	64.4

Feature Category	Feature(s)	No Mortality	Mortality
Demographic Factors	BMI: 25=< x <30	27.9%	32.6%
	Biological Sex: Female	59.0%	51.2%
	Insurance Status: Medicaid only	36.7%	20.9%
	Race/Ethnicity: Non-Hispanic Black	59.8%	53.5%
Utilization Patterns	Number of ED visits in the past year	0.72	0.58
Behavioral Factors	Smoking Status: Never Smoker	72.3%	74.4%
Medication Administration	Covid Vaccination Status: Not vaccinated or incomplete vaccination course	67.7%	76.7%
	Top 5 Medications Documented at Index Encounter	Misc. analgesics: 28.5% Nonsteroidal anti-inflammatory agents: 19.4% 5HT3 receptor antagonists: 13.5% Misc. antivirals: 12.4% Adrenergic bronchodilators: 9.7%	Misc. analgesics: 53.4% Macrolides: 18.6% Adrenergic Bronchodilators: 18.6% Glucocorticoids: 16.3% Misc. Antiemetics: 11.6%
Diagnostic	EKG performed at index encounter	45.0%	69.8%
Encounter Acuity	ESI: 3-Vertical	31.3%	16.3%
	ESI (mean)	3.2	2.7
	Mode of Arrival: Walk-In	84.1%	65.1%
Comorbidities	Top 5 Comorbidities	Obesity: 31.2% Hypertension: 27.3% Asthma: 12.9% Diabetes Type I or II: 11.9% Coronary Artery Disease: 5.7%	Hypertension: 46.5% Obesity: 32.6% Diabetes Type I or II: 32.6% Chronic Kidney Disease: 20.9% Congestive Heart Failure: 18.6%
Vitals (last documented before discharge)	Diastolic Blood Pressure (mean)	77.7	75.7
	Systolic Blood Pressure (mean)	131.2	137.3
	Heart Rate (mean)	82.8	85.3
	Respiratory Rate (mean)	18.0	18.4
	Temperature (mean)	37.0	37.2
	SpO2 (mean)	97.8	96.6
	Ambulatory SpO2 (indicator, % documented during index visit)	21.5%	14.0%
Lab Tests – number of results outside normal range*	BUN (mean)	0.23	0.49
	CO2 (mean)	0.08	0.16
	Creatinine (mean)	0.15	0.40
	PT (mean)	0.01	0.07
	Neutrophil, Percentage (mean)	0.17	0.30
	AST (mean)	0.11	0.28
	Hgb (mean)	0.14	0.33
	BNP (mean)	0.01	0.17
	Platelet (mean)	0.08	0.23
	Glucose Level Random (mean)	0.09	0.37
	Lipase (mean)	0.01	0.09

*Only lab values with the delta between the mean number of results outside of the normal range for hospitalized and non-hospitalized patients within 9 days of the index visit are reported.

Table 9: 9-Day ED return mortality model results

				ROCAUC				F1			
Feature Type	% Return	Number of Encounters	Number of Features	Train	Test			Train	Test		
					Mean	95% CI	STD		Mean	95% CI	STD
Structured EHR Data	0.2%	26454	10464	1	0.805	[0.74 - 0.869]	0.046	0.112	0.015	[0.003 - 0.027]	0.009
NLP: BERT Embeddings	0.2%	23863	3072	1	0.578	[0.484 - 0.672]	0.068	1	0	[0.0 - 0.0]	0
NLP: Bag-of-Words + Logic	0.2%	26454	620	0.923	0.722	[0.678 - 0.767]	0.032	0.0176	0.01	[0.008 - 0.0127]	0.002
All Features	0.2%	26454	14156	1	0.718	[0.57 - 0.866]	0.107	0.997	0	[0.0 - 0.0]	0

Multi-system testing
For the evaluation comparing generalizability to a second health system, the 9-day ED return model had a ROCAUC of 0.659 (95% CI, 0.650 to 0.667) with health system 1 (HS1) test data and a ROCAUC of 0.684 (95% CI, 0.672 to 0.696) with health system 2 (HS2) test data. (Table 10) The model favored recall over precision averaging 0.568 and 0.194 respectively. F1 score was 0.254 and 0.322 for HS1 and HS2 test data, respectively. PRAUC was 0.182 and 0.334 for HS1 and HS2 test data, respectively. Overall, the ED return model performed better across all metrics with HS2 test data.

Table 10: Test performance of 9-day ED return model results on HS1 and HS2 test data

Test data	ROCAUC [95% CI]	Recall [95% CI]	Precision [95% CI]	F1 [95% CI]	PRAUC [95% CI]
HS1	0.659 [0.650, 0.667]	0.446 [0.418, 0.474]	0.178 [0.172, 0.183]	0.254 [0.244, 0.264]	0.182 [0.173, 0.191]
HS2	0.684 [0.672, 0.696]	0.690 [0.668, 0.711]	0.210 [0.203, 0.218]	0.322 [0.312, 0.332]	0.334 [0.316, 0.352]
Average	0.672 [0.661, 0.682]	0.568 [0.543, 0.593]	0.194 [0.188, 0.201]	0.288 [0.278, 0.298]	0.258 [0.245, 0.272]

Increased age and abnormal lymphocyte values, including lymphocytosis and lymphopenia, were noted to have the strongest impact that increased the predicted likelihood of a return in the multi-system model evaluation. Vital signs, specifically higher temperatures and respiratory rates prior to discharge also increased the likelihood of an ED return. In addition, several laboratory values were noted to increase the likelihood of return, including abnormal glucose levels and abnormal platelets. Diabetes as a comorbidity and those diagnosed with pneumonia at the index visit were also identified to increase the predicted likelihood of a return. Male biological sex is also associated with higher return risk, though this may be reflective of known differences by biological sex in comorbidity rates, such as diabetes. Patients who have received monoclonal antibody treatment or have completed a COVID-19 vaccination series, have a decreased likelihood of return. Higher SpO2 values also decreases the likelihood of a return.

Aim 3: A near real-time application found that about 20% of the patients identified would be of concern and this is largely driven by number of prior visits. We found that a leading indicator for ED revisit to be prior history of ED revisits. In addition to past ED encounters, we also found that higher values of the patient's last recorded heart rate, their age at the visit, a higher number of recorded respiratory rate results outside of the normal range, and lower values of SpO2 drove model prediction towards an ED return.

Conclusions
A ML model was trained and tested to predict ED returns amongst COVID-19 patients. The initial predictive model using structured EHR data performed on par with similar predictive models in existing literature that sought to predict ED returns for the general ED population, with previous studies reporting AUCs between 0.60 – 0.883 for models trained on a variety of ED return windows (3-, 7-, 9-, and 30-day returns). Performance did not degrade when applying the SME qualitative concepts to the raw EHR features, providing a level of quantitative validity to the expert-reported clinical factors. NLP feature models performed on par with the baseline model and future results can determine if language used in radiology notes or ED note sections provide more predictive power than EHR data alone. All models performed better when trained on the HIE-enhanced ED return target variable.

Significance
There are several significant aspects to the body of research conducted under this grant in the broader context of understanding the factors associated with ED return risk among COVID-19 patients and its generalizability across different regions and health systems. This research provides important considerations to facilitate safe ED dispositions for individuals with COVID-19, an infectious disease that continues to have significant prevalence in the post-pandemic era.

The findings from the mixed methods concept map development highlight the most salient factors associated with the risk for ED return and morbidity or mortality among returns, which can help narrow down the number of factors considered in risk stratification for ED disposition decisions. For example, existing tools, such as the Emergency Department COVID-19 management tool developed by the American College of Emergency Physicians (ACEP) contains 27 elements that must be manually completed to produce a result and is therefore not readily amenable for use in a fast-paced clinical environment to support point of care decisions.

Our ML model training and testing yielded overall performance metrics comparable to prior literature predicting ED returns for the general population. When comparing model types, the ED return model trained on structured data collected from the EHR outperformed all other models. That being said, the bag-of-words model with features from extracted clinical text displayed promising performance with less model complexity and highlights the utility of NLP. Further, the findings of improved model performance with the HIE-enhanced target variable demonstrate the benefits of HIE data to predict post-discharge events that can occur across a multitude of health systems in a given service area.

Finally, our large-scale, multi-site analysis of emergency department returns among COVID-19 patients demonstrates the generalizability of the ML model developed across two different regional health systems using two different EHRs.

Implications

There are several implications from the body of research conducted under this grant.

First, the findings of this research can inform the development of a screening tool to support ED disposition decisions for individuals with COVID-19. Further exploration that refines the data extraction while continuing to center the predictive power extracted through the distillation of clinical notes is necessary, especially as technology continues to evolve.

Also, this research also shed light on the value of leveraging HIE data to improve model predictive power. HIE data provides more accurate and complete patient health information and utilization records than would otherwise be available in the EHR within a single health system. However, barriers to facilitating the integration of HIE data require mitigation to reduce the practical cost of leveraging HIE information in predictive tools.

Future work could leverage the findings from this research to prospectively optimize model performance as a point-of-care tool that is integrated into clinical workflows to support ED dispositions for COVID-19 patients.

Section V: List of Publications and Products

Peer-reviewed Journal Publications

1. Adams, et al. (Under Review) Utilizing Health Information Exchange to Predict Emergency Department Returns Amongst COVID-19 Patients. *JAMIA.*
2. Galarraga et al. (Under Review) Perceived Factors Associated with Emergency Department Return Risk Among COVID-19 Patients. *Academic Emergency Medicine*.
3. Fong A, et al. (Under Review) Using Machine Learning to Predict ED return amongst COVID-19 patients across two regional health systems. *JAMA Network Open*.

Abstract Presentations

1. Galarraga J, Giovannetti E, Rogovin H, Fitzgerald B, Ndjonko L, & Fong A. (May 2023) Perceived factors associated with emergency department return risk among COVID-19 patients. Abstract presented at the MedStar Health Research Institute Annual Research Symposium. Columbia, MD.
2. Fong A, Adams K, Zhang G, Rogovin H, Giovannetti E, & Galarraga J. (May 2023). Machine-learning prediction of a 9-day return to the emergency department for COVID-19 patients. Abstract presented at the MedStar Health Research Institute Annual Research Symposium. Columbia, MD.
3. Fong A, Adams K, Zhang G, Khairat S, & Galarraga J. (Under Review). Using Machine Learning to Predict ED return amongst COVID-19 patients across two regional health systems. Abstract submitted to the *American College of Emergency Physicians 2024 Annual Scientific Assembly.* Las Vegas, NV.
4. Galarraga J, Giovannetti E, Rogovin H, Fitzgerald B, Ndjonko L, Fong A, & Carlson L. (Under Review) Perceived factors associated with COVID-19 ED returns. Abstract submitted to the *American College of Emergency Physicians 2024 Annual Scientific Assembly.* Las Vegas, NV.
   Adams K, Fong A, Zhang G, & Galarraga J. (Under Review) Abstract submitted to the *American Medical Informatics Association 2024 Annual Symposium.* San Francisco, CA.

Invited Presentation

Conference on Health IT and Analytics (CHITA) - AI and Disparities Panel
This was an invited presentation that described the R21 project, how AI/ML can advance equity in emergency care outcomes, and issues relevant to the application of AI/ML to address health disparities. The project was presented on a panel at the CHITA annual conference in March 2022, sponsored by AHRQ. Its audience included prominent research scholars and policy & practice leaders from more than 40 institutes, with 140 attendees. https://www.rhsmith.umd.edu/centers/chids/chita