Using Real-World Data and Machine Learning to Identify Patients at Highest Risk for Hospitalization Following Respiratory Syncytial Virus Infection

MSR130

Zuzanna Drebert, PhD[1], Julia A O'Rourke, PhD[1], E. Susan Amirian, PhD[1] , Mike Temple, MD[1]
[1]TriNetX, LLC., Cambridge, MA, USA.

BACKGROUND

- Respiratory syncytial virus (RSV) infection is a common cause of respiratory illness in adults, posing a greater risk to patients with chronic medical conditions than to healthy individuals.
- In 2023, the U.S. Food and Drug Administration (FDA) approved two RSV vaccines for older adults and The Centers for Disease Control and Prevention (CDC) recommends vaccination for adults aged 75 years and older and adults aged 60-74 years with chronic medical conditions.

OBJECTIVES

- This work aimed to leverage machine learning (ML) to explore how variables derived from electronic health records (EHR) influence the risk of RSV-related hospitalization.
- We used Explainable AI to understand how these clinical and healthcare utilization characteristics contribute to this risk.

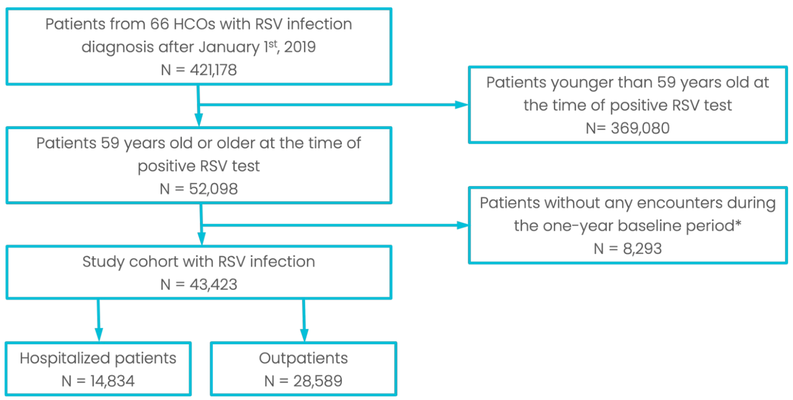

Figure 1. Cohort inclusion and exclusion criteria

METHODS

- This study used the Dataworks-US research network of 66 Healthcare Organizations (HCOs) and 116 million de-identified patients, developed and maintained by TriNetX.
- The cohort included 43,423 adults 59 years old or older with RSV diagnosis or positive lab result between January 1st, 2019, and June 30th, 2024.
- We used Gradient-Boosted Trees (GBT) ML algorithm (80% / 20% training to test split) with 232 demographics, diagnoses, procedures, laboratory, and medications variables selected by the clinical research team to train the model.
- Model performance was evaluated using Receiver Operating Characteristic (ROC), precision, and recall and SHAP (Shapley Additive Explanations) was used to interpret model decisions and to determine relative feature importance.

Table 1. Patient cohort baseline characteristics

CHARACTERISTICS	HOSPITALIZED* N=14,834	OUTPATIENT N=28,589
Age (years)		
Mean (SD)	75.3 (9.63)	73.1 (9.23)
Median [Min, Max]	75.0 [59.0, 92.0]	72.0 [59.0, 93.0]
Sex		
Male	5,831 (39.3%)	10,323 (36.1%)
Female	8,134 (54.8%)	16,467 (57.6%)
Unknown	869 (5.9%)	1,799 (6.3%)
Race		
White	10,380 (70.0%)	21,002 (73.5%)
Black or African American	1,850 (12.5%)	2,770 (9.7%)
Asian	450 (3.0%)	558 (2.0%)
Other	558 (3.8%)	1,304 (4.6%)
Unknown	1,596 (10.8%)	2,955 (10.3%)
US geographic region		
South	5,702 (38.4%)	9,247 (32.3%)
Northeast	5,540 (37.3%)	10,262 (35.9%)
Midwest	1,721 (11.6%)	4,817 (16.8%)
West	1,591 (10.7%)	4,138 (14.5%)
Unknown	280 (1.9%)	125 (0.4%)
RSV diagnosis setting		
ER setting	14,271 (96.2%)	12,090 (42.3%)
Outpatient setting	238 (1.6%)	13,427 (47.0%)
Other setting	23 (0.2%)	842 (2.9%)
Unknown setting	302 (2.0%)	2,230 (7.8%)
Comorbidities		
Any malignancy	3,058 (20.6%)	5,550 (19.4%)
Chronic pulmonary disease	6,023 (40.6%)	8,141 (28.5%)
Congestive heart failure	4,665 (31.4%)	5,130 (17.9%)
Dementia	1,166 (7.9%)	1,440 (5.0%)
Diabetes with chronic complications	3,598 (24.3%)	5,001 (17.5%)
HIV/AIDS	160 (1.1%)	229 (0.8%)
Hemiplegia or paraplegia	341 (2.3%)	478 (1.7%)
Metastatic solid tumor	788 (5.3%)	1,207 (4.2%)
Mild liver disease	1,241 (8.4%)	2,121 (7.4%)
Moderate or severe liver disease	245 (1.7%)	261 (0.9%)
Renal disease	3,459 (23.3%)	4,302 (15.0%)
Rheumatologic disease	758 (5.1%)	1,378 (4.8%)
Asthma	1,882 (12.7%)	3,415 (11.9%)
Obesity	3,112 (21.0%)	5,605 (19.6%)
Lower respiratory tract infections	3,435 (23.2%)	4,868 (17.0%)
Immunocompromised	1,378 (9.3%)	2,561 (9.0%)

*Patients were considered Hospitalized if they had an inpatient encounter that was greater than 24h within 0-1 days from the RSV record. All other cases are considered outpatient.

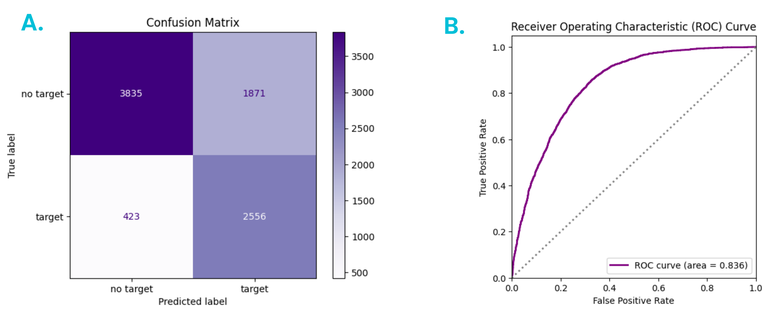

Figure 2. Gradient-Boosted Trees model evaluation: A. Confusion Matrix after probability threshold optimization to 0.279, B. ROC Curve

RESULTS

- Over a third (34.2%) of the patients were admitted to the hospital within one day after the initial RSV diagnosis or positive lab result.
- SHAP analysis revealed that resource utilization patterns were more significant predictors of RSV-related hospitalizations than age and chronic conditions.
- Patients with a history of hospitalizations and lack of outpatient visits were more likely to be hospitalized (variables ranked 1st and 2nd based on SHAP feature importance, with values 0.306 and 0.281, respectively).
- Our trained GBT model achieved a ROC area under curve of 0.836. After optimizing the probability threshold, model's recall improved to 0.858 with a precision of 0.577.

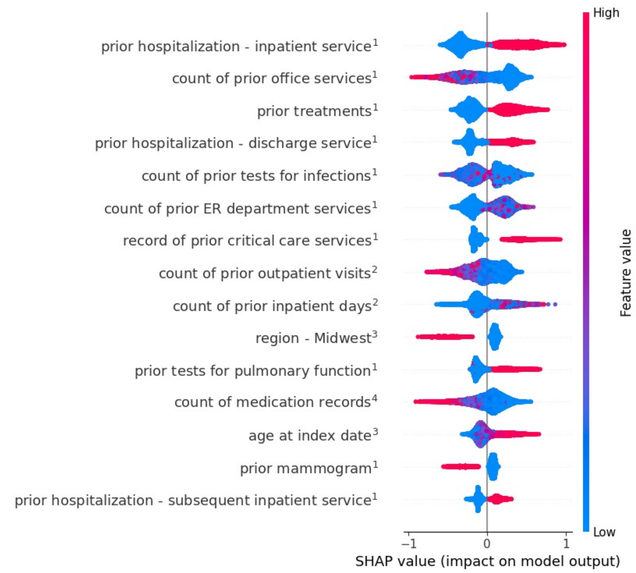

Figure 3. Beeswarm plot with SHAP values for the top 10 features based on SHAP mean feature importance. Features categories: [1]procedures (Restructured BETOS Classification System), [2]encounter type, [3]demographics, [4]medications

CONCLUSION

- Patterns of healthcare utilization appear to be key predictors of RSV-related hospitalization.
- Future work will assess the relevance of these findings for RSV vaccination recommendations.

Download Poster PDF

ISPOR 2025 | Montréal, QC, Canada
Design services were provided by Kelly Marsocci and were funded by TriNetX, LLC.