

Machine Intelligence for Early Targeted Precision Management and Response to Outbreaks of Respiratory Infections

Tiange Zhan, MS; Dev Goyal, MS; John Guttag, PhD; Rudra Mehta, BS; Zahoor Elahi, MS; Zeeshan Syed, PhD; and Mohammed Saeed, MD, PhD

On March 11, 2020, the World Health Organization declared coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), to be a pandemic (ie, a global outbreak occurring on a scale that crosses international boundaries and affects a large number of people). There is much that we do not know about COVID-19. Estimates for lethality range from less than 0.1%¹ to more than 5%.² What we do know is that it is remarkably contagious, with a reproductive number between 2 and 2.5 compared with a little over 1 for influenza.³

While scientists work rapidly to develop therapeutics and vaccines, public health officials have resorted to measures designed to slow the spread of the disease. These efforts to “flatten the curve” aim to preserve capacity and prevent the health system from being overwhelmed by demands for limited resources such as intensive care unit beds, ventilators, and trained care providers. Flattening the curve is targeted toward the population at large. Augmenting these efforts with precise and proactive outreach to higher-risk subpopulations would provide significant value. This subgroup, comprising individuals most vulnerable to severe infections, requires careful and targeted action that goes beyond social distancing and other conventional approaches used for the population at large.

Unfortunately, there are not yet enough data about COVID-19 to draw strong conclusions about who is likely to experience severe sequelae from acquiring the disease. Serious complications associated with COVID-19 are not restricted to elderly or frail individuals; nearly 40% of those hospitalized with COVID-19 in the United States are younger than 54 years.⁴

Using the large amount of historical data about other viral and bacterial respiratory infections, however, it is possible to build useful models for predicting vulnerable individuals who might be at highest risk of severe infections of COVID-19 that require emergent hospitalization. This study explores the precision management of patients at risk of severe respiratory infections—more specifically, personalized and predictive management using machine learning (ML) to provide early targeted management to reduce risk. In particular, this study explores 2 hypotheses centered on the potential for

ABSTRACT

OBJECTIVES: To evaluate the utility of machine learning (ML) for the management of Medicare beneficiaries at risk of severe respiratory infections in community and postacute settings by (1) identifying individuals in a community setting at risk of infections resulting in emergent hospitalization and (2) matching individuals in a postacute setting to skilled nursing facilities (SNFs) that are likely to reduce the risk of infections.

STUDY DESIGN: Retrospective analysis of claims from 2 million Medicare beneficiaries for 2017-2019.

METHODS: In the first analysis, the rate of emergent hospitalization due to respiratory infections was measured among beneficiaries predicted by ML to be at highest risk and compared with the overall average for the population. In the second analysis, the rate of emergent hospitalization due to respiratory infections was compared between beneficiaries who went to an SNF with lower predicted risk of infections using ML and beneficiaries who did not.

RESULTS: In the community setting, beneficiaries predicted to be at highest risk had significantly increased rates of emergency department visits (13-fold) and hospitalizations (18-fold) due to respiratory infections. In the postacute setting, beneficiaries who received care at top-recommended SNFs had a relative reduction of 37% for emergent care and 36% for inpatient hospitalization due to respiratory infection.

CONCLUSIONS: Precision management through personalized and predictive ML offers the opportunity to reduce the burden of outbreaks of respiratory infections. In the community setting, ML can identify vulnerable subpopulations at highest risk of severe infections. In postacute settings, ML can inform patient choices by matching beneficiaries to SNFs likely to reduce future risk.

Am J Manag Care. 2020;26(10):445-448. doi:10.37765/ajmc.2020.88456

TAKEAWAY POINTS

The significant findings of this study include:

- ▶ A machine learning (ML) algorithm accurately predicted risk of 90-day hospitalizations associated with respiratory infections in a large Medicare beneficiary population using only administrative claims data.
- ▶ An ML algorithm recommended skilled nursing facilities (SNFs) associated with lower rates of hospital readmissions or emergency department visits due to respiratory infections during an SNF stay based on each patient’s medical characteristics and historical performance of SNFs in providing care for similar patients.
- ▶ During the current coronavirus disease 2019 (COVID-19) pandemic, ML algorithms have the potential to identify high-risk patients in community and postacute settings who warrant preemptive management to reduce the risk of severe respiratory illness associated with COVID-19.

METHODS

Administrative claims data from roughly 2 million Medicare beneficiaries in community settings between 2017 and 2019 were utilized in this retrospective study. The primary outcome was urgent or emergent inpatient hospital admission with a primary *International Classification of Diseases, Tenth Revision* code related to respiratory infections. A commercially available ML algorithm (HEALTH[at]SCALE Corporation; San Jose, California) was trained on administrative claims data from 2017 and 2018 to predict the risk of respiratory infection-related hospital admission within 90 days.

For training purposes, feature sets and outcomes were constructed from 2017 and 2018 data such that each 90-day outcome window ended prior to December 31, 2018. This algorithm was then applied to 2018 administrative claims data to predict the risk of respiratory infection-related hospital admission in the first quarter of 2019. Data from 2019 were used only for evaluation (ie, used as a clean held-out test set that had no overlap with the training data). The incidence of respiratory infection-related admissions was compared between the top 0.1% and the top 1% of the cohort with the highest predicted risk and the overall population during a 90-day evaluation period.

In the second analysis, administrative claims data from Medicare beneficiaries receiving postacute care at SNFs were utilized. The primary outcome that was measured was the incidence of inpatient hospital readmission or emergency department (ED) visits associated with respiratory infections during the course of SNF stays or within a week of SNF discharge. A commercially available ML algorithm (HEALTH[at]SCALE; San Jose, California) was trained on data from 2017 and 2018 to predict the risk of outcomes for each SNF based on patient medical characteristics at the time of discharge. For training purposes, feature sets and outcomes were constructed from 2017 and 2018 data such that each outcome window ended prior to December 31, 2018. The algorithm was then applied to postacute SNF stays in 2019 (again used as a clean held-out test set that had no overlap with the training data). The rate of respiratory infections that required ED visits or inpatient hospital care was compared between patients who received care at 1 of 3 top-predicted SNFs (as generated by the ML algorithm) and patients who received care at facilities that were not among the top 3 predicted SNFs.

RESULTS

Community Setting

Table 1 shows the characteristics of the Medicare beneficiary population for this analysis. In the community setting, using trailing 12 months of claims data, ML generated a subcohort comprising the top 1% (and also the top 0.1%) of patients estimated to be at greatest risk of severe respiratory infections over the next 90 days.

TABLE 1. Characteristics of Medicare Beneficiary Population

Population size	2 million (approximately)
Age in years, mean (SD)	70.18 (11.93)
Characteristic	Proportion
Age in years	
50-64	14%
65-69	28%
70-74	22%
75-79	15%
80-84	10%
85-89	7%
≥ 90	4%
Gender	
Female	56%
Male	44%
90-day outcomes for all causes	
ED rate	10.90%
Hospitalization rate	5.62%
ED or hospitalization rate	12.17%
90-day outcomes for respiratory infections	
ED rate	0.97%
Hospitalization rate	0.82%
ED or hospitalization rate	1.10%

ED, emergency department.

precision management to reduce the risk of infections for vulnerable individuals: First, precision medicine can be used to predict patients’ personalized risk of adverse outcomes due to respiratory infections in the community setting, and it offers an opportunity to reduce their burden on the health care system through targeted interventions. Second, patient risk of contracting respiratory infections at skilled nursing facilities (SNFs) for postacute care varies based on personalized characteristics, and this risk can be predicted using precision medicine to match patients to facilities that minimize this risk.

TABLE 2. Risk Stratifying Medicare Patients in the (A) Community and (B) Postacute Settings Based on Their Risk of Developing Respiratory Infections

A. Stratifying patients in the community based on their risk of developing respiratory infections					
		Proportion in full population	Proportion in predicted 1% at-risk cohort	Proportion in predicted 0.1% at-risk cohort	Area under the receiver operating characteristic curve
90-day outcomes for respiratory infection	ED rate	0.97%	13.3%	26.4%	0.79
	Hospitalization rate	0.82%	14.6%	32.0%	0.83
	ED or hospitalization rate	1.10%	16.5%	35.3%	0.80
B. Matching patients receiving postacute care to SNFs based on their risk of developing respiratory infections ^a					
		Proportion in top-recommended SNFs	Proportion in non-top-recommended SNFs		
Outcomes within the SNF stay or within a week of SNF discharge	ED rate	5.5%	8.7%		
	Hospitalization rate	3.9%	6.1%		

ED, emergency department; SNF, skilled nursing facility.

^aAll differences in outcomes between top-recommended and non-top-recommended SNFs are statistically significant.

The predictions were compared with the actual outcomes for these patients over this period. Analysis of the claims data for the 90-day evaluation period showed that patients estimated to be at greatest risk were at a 13-fold increased risk of ED visits (13.3% vs 1.0%), an 18-fold increased risk of hospitalization (14.6% vs 0.8%), and a 15-fold increased risk of either ED visits or hospitalizations (16.5% vs 1.1%) due to respiratory infections relative to the population average (Table 2 [A]). The area under the receiver operating characteristic curve was 0.79 or greater for all outcomes (predicting ED visits, hospitalizations, and the composite of the 2 outcomes).

Postacute Setting

In the postacute setting, patients receiving care at SNFs recommended by ML achieved substantial improvements in outcomes relative to those who received care at nonrecommended SNFs. Specifically, patients receiving care at 1 of the top 3 machine intelligence–based personalized SNF recommendations within their local geography had relative reductions of 37% in ED visits (5.5% vs 8.7%) and 36% in inpatient hospitalizations (3.9% vs 6.1%) due to respiratory infections during the course of SNF stays or within a week of SNF discharge (Table 2 [B]). There was a corresponding 26% relative reduction in all-cause ED visits for SNFs recommended by ML (26% vs 35%).

DISCUSSION

In this study, the risk of hospitalizations due to respiratory infections in Medicare beneficiaries in community settings and postacute settings (after hospital discharge) was analyzed. In the community setting, ML accurately classified patients who had the highest risk of hospitalization within 90 days due to respiratory infections. In the postacute setting where Medicare patients were discharged to SNFs, patients who went to SNFs recommended by ML had significantly lower rates of hospital readmissions and ED visits associated with respiratory infections compared with patients who went to SNFs that were not recommended for them.

The current pandemic associated with novel coronavirus–infected pneumonia (NCIP) has swept across continents and afflicted more than 13 million people worldwide and has been attributed to more than 570,000 deaths due to highly efficient human-to-human transmission.³ Morbidity and mortality have been particularly high in elderly patients and those with chronic comorbidities due to severe respiratory complications associated with NCIP.^{3,5} It may be possible to improve outcomes and reduce the burden on the health care system by identifying vulnerable patient populations and utilizing preemptive and preventive outreach to reduce their risk.

Given that SARS-CoV-2 is a relatively new pathogen associated with severe respiratory infections, the performance of an ML algorithm in identifying high-risk or “vulnerable” patients in the community and postacute settings was evaluated. Based on the current epidemiologic and outcomes data regarding SARS-CoV-2, patients who have been vulnerable to severe complications from viral influenza, community-acquired pneumonia (CAP), and health care–associated pneumonia would likely be vulnerable to SARS-CoV-2, as well.^{3,6-10} In a study examining risk factors for CAP, it was found that 70% of hospitalized patients were 50 years and older and 78% had 1 or more chronic comorbidities including congestive heart failure, chronic obstructive pulmonary disease, diabetes, or immunosuppression.¹¹ In a recently published systematic review of several risk models for SARS-CoV-2, similar risk factors related to advanced age and chronic comorbidities were found to be associated with severe SARS-CoV-2 infections.¹² Thus, in the absence of large-scale SARS-CoV-2 patient data for training ML models, use of prior patient data associated with patients hospitalized for respiratory infections may suffice.

Among Medicare beneficiaries in community settings, patients who were classified by ML to be in the top 1% of risk for respiratory infection–related hospitalization within 90 days had an 18-fold higher risk than the baseline Medicare population. In the current COVID-19 pandemic, a similar ML algorithm could be utilized to identify patients who are at high risk of developing severe respiratory illnesses due to COVID-19. Targeted programs aimed at these high-risk patients during

TRENDS FROM THE FIELD

the COVID-19 pandemic may include reinforcement of guidelines about social distancing, contact precautions, symptom education, home delivery of medications, house calls/televisits, portal engagement, and other preventive outreach to mitigate risk in this cohort.

Besides community patients, patients being discharged from hospitals for postacute care are at severely elevated risk of experiencing substantial mortality or morbidity as a result of infection. Prior research has shown differential performance of SNFs in reducing hospital readmissions and mortality for Medicare beneficiaries in the postacute setting.¹³ Thus, the choice of SNF destination may influence outcomes in vulnerable Medicare patients in the postacute setting. An ML algorithm was utilized for selecting SNF destinations, and patients who went to SNFs that were recommended by the ML algorithm were shown to have lower rates of ED visits and hospitalizations due to respiratory infections. Such personalized recommendations for matching patients to facilities are especially pertinent given the requirements currently in place for hospitals to provide patients with information about postacute care provider choices. Personalized recommendations for the choice of SNF driven by historic data are intended to augment the decision-making process for caregivers and patients while preserving choice and focusing on optimizing long-term outcomes. Furthermore, we believe that personalized, ranked recommendations for facilities might be more actionable than statistics about facilities derived from the general population. In the current pandemic, use of ML tools for clinical decision support in the postacute setting may similarly help in preventing SARS-CoV-2–related severe illness.

CONCLUSIONS

Future work will explore the use of recent data collected from SARS-CoV-2–affected patients to update our model. Limitations of this study include using only administrative claims data, which may not fully characterize patient frailty and medical problems as well as the clinical data available in electronic health records. Medicare beneficiaries are typically older; thus, the cohort in this study does not reflect predictive performance of the model on younger patients who are also susceptible to respiratory infections. Risk of respiratory infections is also influenced by exposure risk due to disease prevalence in the community¹⁴ and other socioeconomic factors such as population density, patient contacts, and air quality index that were not evaluated in this study. The proposed matching of patients to facilities for postacute care was not evaluated in prospective observational causal studies or randomized clinical trials and would benefit from follow-up investigation in studies of this nature. ■

Acknowledgments

Tianghe Zhan, MS, and Dev Goyal, MS, contributed equally to this work and are listed as co–first authors.

Author Affiliations: HEALTH[at]SCALE Corporation (TZ, DG, JG, RM, ZE, ZS, MS), San Jose, CA; Massachusetts Institute of Technology (JG), Cambridge, MA; University of Michigan (MS), Ann Arbor, MI.

Source of Funding: HEALTH[at]SCALE Corporation.

Author Disclosures: Ms Zhan, Mr Goyal, and Mr Mehta are employees of and stock owners in HEALTH[at]SCALE Corporation, which developed the machine learning algorithms described in the paper. Dr Guttag is a board member and consultant for HEALTH[at]SCALE, owns stock in HEALTH[at]SCALE, has attended the Becker's Hospital Review conference, and reports patents pending filed by HEALTH[at]SCALE and patents received licensed by HEALTH[at]SCALE. Mr Elahi is an employee of HEALTH[at]SCALE and owns stock in HEALTH[at]SCALE. Drs Syed and Saeed are board members and employees of HEALTH[at]SCALE, report patents pending from HEALTH[at]SCALE, and own stock in HEALTH[at]SCALE.

Authorship Information: Concept and design (JG, ZE, ZS, MS); acquisition of data (TZ, DG, RM); analysis and interpretation of data (TZ, DG, RM, MS); drafting of the manuscript (TZ, DG, JG, ZE, ZS, MS); critical revision of the manuscript for important intellectual content (TZ, DG, JG, ZS, MS); statistical analysis (TZ, DG); administrative, technical, or logistic support (RM); and supervision (ZE, ZS, MS).

Address Correspondence to: Mohammed Saeed, MD, PhD, University of Michigan, 1500 E Medical Center Dr, Ann Arbor, MI 48109. Email: msaeed@med.umich.edu.

REFERENCES

- Ioannidis JPA. A fiasco in the making? as the coronavirus pandemic takes hold, we are making decisions without reliable data. *STAT*. March 17, 2020. Accessed March 26, 2020. <https://www.statnews.com/2020/03/17/a-fiasco-in-the-making-as-the-coronavirus-pandemic-takes-hold-we-are-making-decisions-without-reliable-data/>
- Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. *Lancet Infect Dis*. 2020;20(7):773. doi:10.1016/S1473-3099(20)30195-X
- Coronavirus disease 2019 (COVID-19): situation report, 51. World Health Organization. March 11, 2020. Accessed March 12, 2020. <https://apps.who.int/iris/bitstream/handle/10665/331475/nCoVsitrep11Mar2020-eng.pdf>
- Mansell W, Flaherty A. Nearly 40% of those hospitalized with coronavirus are younger than 54: CDC. ABC News. March 19, 2020. Accessed March 26, 2020. <https://abcnews.go.com/US/40-hospitalized-coronavirus-younger-54-cdc/story?id=69681304>
- Li Q, Guan X, Peng W, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*. 2020;382(13):1199-1207. doi:10.1056/NEJMoa2001316
- Bont J, Hak E, Hoes AW, Schipper M, Schellevis FG, Verheij TJM. A prediction rule for elderly primary-care patients with lower respiratory tract infections. *Eur Respir J*. 2007;29(5):969-975. doi:10.1183/09031936.00129706
- Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med*. 1997;336(4):243-250. doi:10.1056/NEJM199701233360402
- Chalmers JD, Singanayagam A, Akram AR, et al. Severity assessment tools for predicting mortality in hospitalised patients with community-acquired pneumonia: systematic review and meta-analysis. *Thorax*. 2010;65(10):878-883. doi:10.1136/thx.2009.133280
- Millett ER, De Stavola BL, Quint JK, Smeeth L, Thomas SL. Risk factors for hospital admission in the 28 days following a community-acquired pneumonia diagnosis in older adults, and their contribution to increasing hospitalisation rates over time: a cohort study. *BMJ Open*. 2015;5(12):e008737. doi:10.1136/bmjopen-2015-008737
- Jain S, Self WH, Wunderink RG, et al: CDC EPIC Study Team. Community-acquired pneumonia requiring hospitalization among US adults. *N Engl J Med*. 2015;373(5):415-427. doi:10.1056/NEJMoa1500245
- Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 2020;369:m1328. doi:10.1136/bmj.m1328
- Neuman MD, Wirtalla C, Werner RM. Association between skilled nursing facility quality indicators and hospital readmissions. *JAMA*. 2014;312(15):1542-1551. doi:10.1001/jama.2014.13513
- Barnes R, Blyth CC, de Klerk N, et al. Geographical disparities in emergency department presentations for acute respiratory infections and risk factors for presenting: a population-based cohort study of Western Australian children. *BMJ Open*. 2019;9(2):e025360. doi:10.1136/bmjopen-2018-025360

Visit ajmc.com/link/88456 to download PDF