

RESEARCH

Open Access



# Evaluating methods for risk prediction of Covid-19 mortality in nursing home residents before and after vaccine availability: a retrospective cohort study

Komal Aryal<sup>1,2\*</sup>, Fabrice I. Mowbray<sup>3</sup>, Anna Miroshnychenko<sup>1</sup>, Ryan P. Strum<sup>1</sup>, Daryl Dash<sup>1</sup>, Michael P. Hillmer<sup>4,5</sup>, Kamil Malikov<sup>5</sup>, Andrew P. Costa<sup>1,2</sup> and Aaron Jones<sup>1,2</sup>

## Abstract

**Background** SARS-CoV-2 vaccines are effective in reducing hospitalization, COVID-19 symptoms, and COVID-19 mortality for nursing home (NH) residents. We sought to compare the accuracy of various machine learning models, examine changes to model performance, and identify resident characteristics that have the strongest associations with 30-day COVID-19 mortality, before and after vaccine availability.

**Methods** We conducted a population-based retrospective cohort study analyzing data from all NH facilities across Ontario, Canada. We included all residents diagnosed with SARS-CoV-2 and living in NHs between March 2020 and July 2021. We employed five machine learning algorithms to predict COVID-19 mortality, including logistic regression, LASSO regression, classification and regression trees (CART), random forests, and gradient boosted trees. The discriminative performance of the models was evaluated using the area under the receiver operating characteristic curve (AUC) for each model using 10-fold cross-validation. Model calibration was determined through evaluation of calibration slopes. Variable importance was calculated by repeatedly and randomly permutating the values of each predictor in the dataset and re-evaluating the model's performance.

**Results** A total of 14,977 NH residents and 20 resident characteristics were included in the model. The cross-validated AUCs were similar across algorithms and ranged from 0.64 to 0.67. Gradient boosted trees and logistic regression had an AUC of 0.67 pre- and post-vaccine availability. CART had the lowest discrimination ability with an AUC of 0.64 pre-vaccine availability, and 0.65 post-vaccine availability. The most influential resident characteristics, irrespective of vaccine availability, included advanced age ( $\geq 75$  years), health instability, functional and cognitive status, sex (male), and polypharmacy.

**Conclusions** The predictive accuracy and discrimination exhibited by all five examined machine learning algorithms were similar. Both logistic regression and gradient boosted trees exhibit comparable performance and display slight superiority over other machine learning algorithms. We observed consistent model performance both before and

\*Correspondence:

Komal Aryal  
aryalk@mcmaster.ca

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

after vaccine availability. The influence of resident characteristics on COVID-19 mortality remained consistent across time periods, suggesting that changes to pre-vaccination screening practices for high-risk individuals are effective in the post-vaccination era.

**Keywords** Long-term care, Nursing home, Older adults, COVID-19, Prediction modeling, Cohort study, Machine learning, Electronic Health Record

## Introduction

The COVID-19 pandemic led to an exponential surge in the number of deaths within nursing homes (NH) [1–4]. In 2020, 63% of all deaths in the NH were attributed to COVID-19 among NH resident deaths in Canada [5]. NH residents have been disproportionately affected by COVID-19 illness due to their complex health and physical care needs, coupled with increasing frailty [6]. SARS-CoV-2 vaccines have been effective in reducing hospitalization, COVID-19 symptoms, and mortality for NH residents, and NH residents were prioritized during vaccine rollout [7–11]. Nearly 80% of NH residents had received one dose of a SARS-CoV-2 vaccine in Canada by the end of January 2021 [12]. Even after vaccination however, residents with high-risk profiles can still experience poor outcomes from COVID-19, including death [7]. Numerous resident characteristics (e.g., age, gender, cognitive status, and physical functioning) have been examined as prognostic factors for COVID-19 mortality [13, 14]. However, little is known about how the predictability of COVID-19 mortality changed due to the vaccine rollout.

Regression-based and tree-based machine learning models have been widely used in health and health services research. Advanced machine learning algorithms demonstrate remarkable capability in identifying high-risk subpopulations, particularly when predictors exhibit intricate interaction effects [15]. As a result, these methods have gained considerable popularity in research involving complex and vulnerable populations, such as NH residents. However, a unanimous consensus on the most suitable method to discriminate outcomes remains elusive, primarily due to the susceptibility of tree-based models to overfitting, which compromises the model's generalizability [16–19]. Accurate mortality prediction at the individual NH resident level could greatly benefit healthcare professionals in prioritizing medical care and enabling efficient resource planning.

Our study aimed to utilize different machine learning methods to compare COVID-19 mortality prognostication in NH residents before and after vaccine availability. Our objectives were to establish the accuracy of various machine learning models, examine changes to model performance, and identify resident characteristics that have the strongest associations with 30-day COVID-19 mortality, before and after availability.

## Methods

### Study design

We conducted a population-based retrospective cohort study analyzing data from all NH facilities across Ontario, Canada. This study was reviewed and approved by the Hamilton Integrated Research Ethics Board (HiREB # 10,959-C). To ensure accurate reporting, we followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement guidelines for this cohort study.

### Data sources

Four population-level health administrative databases were examined. The Continuing Care Reporting System (CCRS) is a data repository of clinical assessments that are completed for each NH resident using the Resident Assessment Instrument - Minimum Data Set (RAI-MDS) 2.0 [20]. Residents receive a RAI-MDS assessment upon admission into the facility and then every three months thereafter. The CCRS also records changes to resident status such as discharges from NH, death died in the facility, or during a hold-bed period when a resident is temporarily elsewhere such as the hospital. The Ontario Laboratories Information System (OLIS) records lab test orders and results from hospitals, community labs and public health labs, including those for COVID-19, for which we used the specimen collection date. The Ontario Integrated Public Health Information System (IPHIS) documents public health cases, including COVID-19, for which we used the episode start date, and records dates of death if the resident died during the COVID-19 episode. We used the Discharge Abstract Database (DAD) to identify all those who were admitted to the hospital and recorded as deceased during their stay. Datasets used were linked and deidentified by the Ontario Ministry of Health (MOH).

### Study participants

We included all residents in our cohort who were admitted into the NH and stayed at an NH for one or more days between March 7, 2020, and July 31, 2021. We excluded residents if they had no complete RAI-MDS 2.0 assessment or resided outside of Ontario. For residents who were admitted to an NH more than once, we utilized the first assessment per time period to avoid correlated data among residents. Our study only utilized assessments

completed on or before the positive SARS-CoV-2 infection date [21].

### Outcome measure & exposure periods

Our primary outcome was 30-day mortality following a laboratory-verified positive SARS-CoV-2 test. We examined positive cases during two time periods. The first time period lasted from the start of the pandemic to the beginning of vaccine availability (March 7, 2020, to December 31, 2020) which was selected based on the closest date to the first positive SARS-CoV-2 infection in NH and the first reported outbreak [22]. The second time period lasted from vaccine availability until the end of the study period (January 1, 2021, to July 31, 2021). We selected January 1, 2021 as our second time period because it approximates the start of vaccine rollout in the NH [23].

Residents in both time periods were followed for 30 days to measure the primary outcome. If residents tested positive for SARS-CoV-2 more than once in each study period, we included only the first instance to avoid correlated data. Patients who tested positive during the first 30 days of the second time period were only included if there was no documentation of a positive COVID-19 test in the previous 30 days.

### Resident characteristics

Resident characteristics for inclusion in the machine learning models were selected based on their availability in our data sources, clinical expertise and prior literature [3, 24–28]. Resident characteristics came from the completed MDS 2.0 assessments. These characteristics include demographic, clinical, and social characteristics reported at the NH facility that are expected to be associated with COVID-19 mortality. The methods employed for resident character selection in the study were not statistically driven and are described in detail in our previous work [29].

### Statistical analysis

Descriptive statistics were reported using measures of frequency and central tendency to compare residents who died due to COVID-19 both before and after vaccine availability. Variable selection was performed a priori using both theoretical and clinical methods based on existing literature [29]. We used five machine learning algorithms to predict COVID-19 death. These included logistic regression, LASSO logistic regression, classification and regression tree (CART), random forests, and gradient boosted trees (GBT). Data were screened for the presence and pattern of missingness. Only five (0.03%) cases of missing data were present for one predictor variable and these cases were deleted within each analysis.

We performed hyperparameter tuning for each non-logistic regression model to determine the optimal set of parameters for each machine learning model. Tuning was performed using 10-fold cross-validation with a 1000-iteration random grid search over the parameter space, selecting the parameters that maximized the area under the receiver operating characteristic curve (AUC). All tuning and testing were performed independently for the two vaccination periods. Final performance was determined by calculating the AUC for each model using an independent and identical 10-fold cross-validation and by computing F1 scores and balanced accuracy. Permutation methods were used to calculate variable importance by repeating ( $n=50$ ) randomly permuted values of each predictor in the dataset and re-evaluating the model's performance, which was not nested inside the cross-validation [30]. The average difference in the AUC was used to measure variable importance, with negative values of greater magnitude indicating more importance. Model calibration was assessed visually using calibration plots. Data were managed and analyzed using R-Studio 4.2.2.

### Sensitivity analysis

We conducted a sensitivity analysis by excluding January 2021, during which time the first dose was being rolled-out to NH residents. We recalculated performance statistics for logistic regression and LASSO regression. Since fewer infections occurred between February 2021 to July 2021, complex machine learning models including random forests, CART, and GBT were not conducted in the sensitivity analysis.

### Results

There were 14,977 NH residents who tested positive for a COVID-19 infection during the study period. A total of 11,291 NH residents tested positive for COVID-19 before vaccine availability and 3686 NH residents tested positive after vaccine availability. The median time (25th–75th percentiles) from the MDS 2.0 assessment date to the COVID-19 date was 44 days (22–67).

Table 1 displays a comprehensive list of resident characteristics for NH residents who were diagnosed with COVID-19 before and after vaccine availability. Most residents were female (65.8%; 67.6%), 65+ years (92.6%; 93.6%), and were diagnosed with dementia (52.8%; 52.7%) in the before and after vaccine availability groups, respectively. There were 2937 (26.0%) NH residents who died before vaccine availability and 727 (19.7%) NH residents who died after vaccine availability.

### Model performance

Table 2 displays the final model performance from the five machine learning models. The cross-validated AUC

**Table 1** Descriptive Characteristics of Nursing Home Residents who were Diagnosed with COVID-19 Infection

Resident Characteristic	Before Vaccine Availability Mar 7, 2020 – Dec 31, 2020 (n = 11,291) n(%)	After Vaccine Availability Jan 1, 2021– July 31, 2021 (n = 3686) n(%)
<b>Age Group</b> , years		
<64	832 (7.4)	235 (6.4)
65–74	1466 (13.0)	481 (13.0)
75–84	3134 (27.8)	994 (27.0)
85+	5859 (51.9)	1976 (53.6)
<b>Sex</b>		
Male	3857 (34.2)	1195 (32.4)
Female	7434 (65.8)	2491 (67.6)
<b>Comorbidities</b>		
Cardiovascular Diseases	2771 (24.5)	898 (24.4)
Congestive Heart Failure	429 (3.8)	183 (5.0)
Diabetes	3431 (30.4)	1128 (30.6)
Respiratory Disease <sup>a</sup>	935 (8.3)	418 (11.3)
Renal Failure	399 (3.5)	140 (3.8)
Dementia	5958 (52.8)	1942 (52.7)
Depression	3576 (31.7)	1229 (33.3)
Anxiety	462 (4.1)	177 (4.8)
Cancer	320 (2.8)	120 (3.2)
Fever and Headache	353 (3.1)	120 (3.2)
<b>Polypharmacy</b>		
Taking 5 or more medications/daily	10,417 (92.3)	3444 (93.4)
Taking less than 5 medications/daily	874 (7.7)	242 (6.6)
<b>Nutrition Risk<sup>b</sup></b>		
	3955 (35.0)	1504 (40.8)
<b>Daily Decision Making</b>		
Independent	914 (8.1)	436 (11.8)
Modified Independence	2283 (20.2)	780 (21.2)
Moderately Impaired	5679 (50.3)	1775 (48.2)
Severely Impaired	2410 (21.3)	695 (18.9)
<b>Activities of Daily Living</b>		
Independent or Supervision Required	501 (4.4)	170 (4.6)
Limited	734 (6.5)	226 (6.1)
Extensive	3046 (27.0)	993 (27.0)
Maximal	3014 (27.7)	921 (25.0)
Dependence	2872 (25.4)	990 (26.8)
Total Dependence	1123 (9.9)	386 (10.5)
<b>Cognitive Performance Scale (CPS)</b>		
Intact	1651 (14.6)	658 (17.8)
Moderate	7224 (64.0)	2333 (63.3)
Severe	2415 (21.4)	695 (18.9)
<b>Pain</b>		
Less than Daily	1742 (15.4)	671 (18.2)
Daily but Not Severe	279 (2.5)	134 (3.6)
Daily and Severe	56 (0.5)	22 (0.6)

<sup>a</sup>COPD, Emphysema, Asthma, & Dyspnea. <sup>b</sup>Decreased appetite, weight loss, & dehydration

performance ranged from 0.64 to 0.67, irrespective of vaccination period. Across all models, GBT had the highest discrimination ability with an AUC of 0.67 for both the before and after vaccination periods, although

the discriminative accuracy was not significantly different from the other machine learning models, excluding CART ( $p < .05$ ). Our CART model had the lowest discrimination ability both before and after the vaccination

**Table 2** Final Model Performance of Logistic and Machine Learning Models Before and After Vaccine Availability

Model Type	Before Vaccine Availability Mar 7, 2020 – Dec 31, 2020 ( <i>n</i> = 11,291)	After Vaccine Availability Jan 1, 2021– July 31, 2021 ( <i>n</i> = 3686)
Logistic Regression	0.670	0.670
Lasso Regression	0.664	0.664
Random Forest	0.662	0.663
Classification and Regression Tree (CART)	0.641	0.646
Gradient Boosted Trees	0.672	0.670

periods compared to logistic regression, random forests, and gradient-boosted trees. F1 scores and balanced accuracy are reported in Appendix A and model calibration are presented in Appendix B. The range of model parameters are reported in Appendix C.

### Associations with COVID-19 mortality

Approximately half (11/20) of the resident characteristics contributed to the performance of all five models (Appendix D). On average, the five most influential resident characteristics both before and after vaccine availability were being 85 years older, being aged 75–84, being male, deteriorating ADLs, and having a high score on the CHES scale (i.e., health instability) (Fig. 1). The least influential resident characteristics across all five models both before and after vaccine availability, having a headache, having a fever, experiencing anxiety, having cancer, congestive heart failure, and having respiratory disease. Overall, there was little difference in the variable importance before and after vaccine availability and between the five models.

### Sensitivity analysis

There were significantly fewer cases in the February 1, 2021, to July 31, 2021 (*n*=828) than January 1, 2021 to July 31, 2021 (*n*=3686). Analysis showed similar but slightly higher AUC values for logistic regression before (0.68) and after (0.69) vaccine availability. Similarly, the AUC values for LASSO regression before (0.66) and after (0.68) were slightly higher after vaccine availability.

### Discussion

We used one statistical model and four machine learning models to examine 30-day COVID-19 mortality among NH residents before and after vaccine availability in Ontario, Canada. All models exhibited similar predictive accuracy both between models and across the COVID-19 vaccine availability time periods. Approximately half of resident characteristics were informative in identifying residents at high-risk of COVID-19 mortality. These factors remained consistent regardless of vaccine availability in all models.

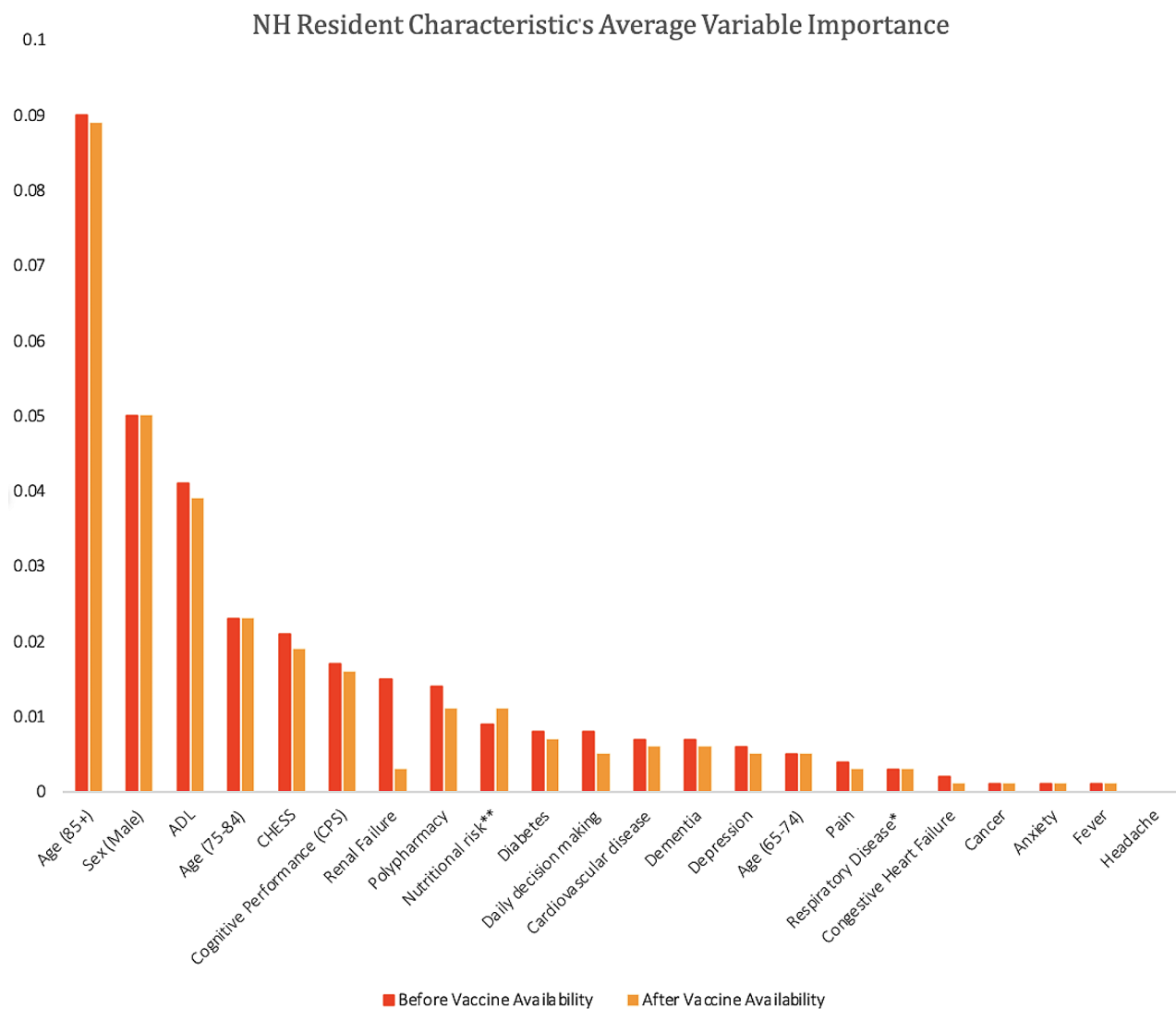
Our study highlights a ceiling effect on the discriminative ability of machine learning algorithms when using

routinely collected administrative data compared to the statistical logistic regression model. While GBT models can accommodate complex patterns within the data, they are computationally complex, and their “black box” nature makes them less appealing to clinical audiences. Prior works have demonstrated similar discriminative accuracies between GBT and logistic regression in complex older adults [30–32]. Further, our study contributes evidence to prior work demonstrating that the use of complex supervised machine learning algorithms is unlikely to out-perform standard regression models using highly structured data [33]. We conducted a comparative analysis of the discriminability of machine learning methods before and after COVID-19 vaccine availability. We found no discernible difference in the performance of these models based on vaccine availability.

Previous studies have reported higher AUC values when predicting 30-day COVID-19 mortality [34]. However, many of these studies focused on 30-day COVID-19 mortality within broad populations, in which age is a highly discriminative predictor of mortality, a consistent result of our study. For example, a study by Hippisley-Cox et al., [26] focused on a broad population of all adults in England and did not assess the risk of COVID-19 mortality for the NH population. In contrast, our study specifically aims to predict COVID-19 mortality among older, frail nursing home residents. The limited heterogeneity in our samples makes it more challenging to discern individuals who are at a higher risk of death.

We sought to evaluate whether the significance and magnitude of the resident characteristics in these models differed between vaccine availability periods. The important resident characteristics in our model were older age, male sex, and deteriorating ADL status with age being the most influential. However, our results indicate that there is little difference in resident characteristics influencing COVID-19 mortality based on vaccine availability. NH resident characteristics alone were not sufficiently able to determine which residents were at greatest risk of COVID-19 mortality, as evidenced by their relatively weak AUC values, irrespective of time period.

From our research, it is evident that pre-vaccination prognostic scores and models are still informative of post-vaccination scores and models could be effective



**Fig. 1** The Average Inverse Variable Importance for all Resident Characteristics Before and After Vaccine Availability. \*Respiratory Disease: Chronic Obstructive Pulmonary Disease, Emphysema, Asthma, & Dyspnea \*\*Nutrition Risk: Decreased appetite, weight loss, & dehydration

when employed post vaccination rollout. Existing practices to identify residents at high-risk of COVID-19 death likely do not need to be adjusted. This finding can help determine future care plans for both vaccinated and unvaccinated NH populations. Future studies should determine if COVID-19 risk factors remain stable for older individuals living in congregate care settings such as retirement homes.

#### Limitations

We leveraged a population-level database of all NH residents across Ontario and reported on a wide array of resident factors and geriatric syndromes known to be prognostic of mortality post-SARS-CoV-2 infection. However, we were limited to secondary data collected

across databases. Undocumented resident characteristics, such as ethnicity or race, may influence mortality but are not recorded in the RAI-MDS 2.0. Our databases did not capture the accurate date or type of vaccine received by NH residents and thus were unable to stratify based on actual vaccination type. These predictors may have been informative in predicting COVID-19 mortality, but we were unable to include them in our model. Our analysis began during the early stages of COVID-19 and some residents may not have had a COVID-19 test before dying results in some COVID-19 deaths may not being captured. The discriminative accuracy of statistical models was *fair* despite having a panel of prognostic factors known to influence 30-day COVID-19 mortality. However, the use of prognostic models with this level

of discriminative ability in population-level research is common, considering the difficulties of predicting within a complex-adaptive system [35, 36].

### Conclusions and implications

Our study determined that all statistical and machine learning algorithms examined displayed similar predictive accuracy. This suggests that there would be no benefit in choosing a more complex tree-based model over standard regression for these data sources. Overall, the performance of the models did not differ before and after vaccine availability, indicating that vaccine uptake did not change COVID-19 mortality prognostication. Resident characteristics influencing 30-day COVID-19 mortality are similar both before and after vaccine availability. The stability of risk factors and performance between vaccination periods suggests that models generated to predict COVID-19 mortality pre-vaccination are valid for use in the post-vaccination era.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-024-02189-3>.

Supplementary Material 1

### Acknowledgments

Thank you to Dr. Jeff Poss for assisting with data acquisition.

### Author contributions

All authors have read and approved the submission of this manuscript. Komal Aryal and Dr. Aaron Jones contributed to the study concept and design. Komal Aryal, Dr. Andrew Costa, Dr. Aaron Jones, Dr. Kamil Malikov, and Dr. Michael P. Hillmer contributed to data acquisition. Dr. Fabrice I. Mowbray and Ryan P. Strum contributed to the clinical conceptualization of the analysis. Komal Aryal, Anna Miroshnychenko and Darly Dash contributed to the descriptive analysis of the manuscript. Komal Aryal contributed to the data analysis and interpretation. Komal Aryal drafted the manuscript. All authors contributed to critical revisions of the manuscript for intellectual content and approved the final version to be published. Drs. Andrew Costa and Aaron Jones provided study supervision.

### Funding

This work was supported in part by grants from the Canadian COVID-19 Immunity Task Force and Public Health Agency of Canada (PHAC) awarded to APC and DMB (2021-HQ-000138). APC is supported by the Schlegel Chair in Clinical Epidemiology and Aging at McMaster University. DD holds a doctoral scholarship through the Canadian Institutes of Health Research (CIHR) (grant #FBD-181577). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

### Data availability

Data access is governed separately by the Ontario Personal Health Information Protection Act and held securely at McMaster University. Analytic coding is available upon request from the authors with appropriate approvals to protect the security of source the data architecture.

### Declarations

#### Ethics approval and consent to participate

This study was a secondary analysis of data from the Canadian Longitudinal Study on Aging (CLSA). This study was approved by the Hamilton Integrated

Research Ethics Board (HiREB #10959-C). This study was conducted in accordance with the Declaration of Helsinki and all participants provided informed consent.

#### Consent for publication

Not applicable in this section.

#### Competing interests

The authors have no competing interests to declare.

#### Author details

<sup>1</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada

<sup>2</sup>ICES, Hamilton, ON, Canada

<sup>3</sup>College of Nursing, Michigan State University, East Lansing, MI, USA

<sup>4</sup>Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada

<sup>5</sup>Capacity Planning and Analytics, Ontario Ministry of Health, Toronto, Canada

Received: 4 August 2023 / Accepted: 22 February 2024

Published online: 27 March 2024

### References

- Shen K, Loomer L, Abrams H, Grabowski DC, Gandhi A. Estimates of COVID-19 cases and deaths among nursing home residents not reported in Federal Data. *JAMA Netw Open*. 2021;4:e2122885.
- Kaiser Family Foundation. State data and policy actions to address coronavirus. Published November 18 AN, 2022. <https://www.kff.org/health-costs/issue-brief/state-data-and-policy-actions-to-address-coronavirus/>.
- Jordan RE, Adab P, Cheng KK. Covid-19: risk factors for severe disease and death. *BMJ*. 2020;368:m1198.
- Zhao HL, Huang YM, Huang Y. Mortality in older patients with COVID-19. *J Am Geriatr Soc*. 2020;68:1685–7.
- <https://health-infobase.canada.ca/covid-19/current-situation.html> GoCC-euCsRftGoCw.
- De Smet R, Mellaerts B, Vandewinckele H, Lybeert P, Frans E, Ombelet S, et al. Frailty and Mortality in hospitalized older adults with COVID-19: Retrospective Observational Study. *J Am Med Dir Assoc*. 2020;21:928–32. e1.
- Lafuente-Lafuente C, Rainone A, Guerin O, Drunat O, Jeandel C, Hanon O, et al. COVID-19 outbreaks in nursing homes despite full vaccination with BNT162b2 of a majority of residents. *Gerontology*. 2022;68:1384–92.
- White EM, Yang X, Blackman C, Feifer RA, Gravenstein S, Mor V. Incident SARS-CoV-2 infection among mRNA-Vaccinated and unvaccinated nursing home residents. *N Engl J Med*. 2021;385:474–6.
- Wouters F, van Loon AM, Rutten JJS, Smalbrugge M, Hertogh C, Joling KJ. Risk of death in nursing home residents after COVID-19 vaccination. *J Am Med Dir Assoc*. 2022;23:1750–3. e2.
- Nilsson L, Andersson C, Kastbom L, Sjudahl R. Association between vaccination and preventive routines on COVID-19-related mortality in nursing home facilities: a population-based systematic retrospective chart review. *Prim Health Care Res Dev*. 2022;23:e75.
- McConeghy KW, Bardenheier B, Huang AW, White EM, Feifer RA, Blackman C, et al. Infections, hospitalizations, and deaths among US nursing home residents with vs without a SARS-CoV-2 Vaccine Booster. *JAMA Netw Open*. 2022;5:e2245417.
- Detsky AS, Bogoch II. COVID-19 in Canada: experience and response to waves 2 and 3. *JAMA*. 2021;326:1145–6.
- Panagiotou OA, Kosar CM, White EM, Bantis LE, Yang X, Santostefano CM, et al. Risk factors Associated with all-cause 30-Day mortality in nursing home residents with COVID-19. *JAMA Intern Med*. 2021;181:439–48.
- Mehta HB, Li S, Goodwin JS. Risk factors Associated with SARS-CoV-2 infections, hospitalization, and Mortality among US nursing home residents. *JAMA Netw Open*. 2021;4:e216315.
- Lo-Ciganic WH, Huang JL, Zhang HH, Weiss JC, Wu Y, Kwok CK, et al. Evaluation of machine-learning algorithms for Predicting Opioid Overdose Risk among Medicare beneficiaries with opioid prescriptions. *JAMA Netw Open*. 2019;2:e190968.

16. Boyce RD, Kravchenko OV, Perera S, Karp JF, Kane-Gill SL, Reynolds CF, et al. Falls prediction using the nursing home minimum dataset. *J Am Med Inf Assoc.* 2022;29:1497–507.
17. Austin PC. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. *Stat Med.* 2007;26:2937–57.
18. Heldt FS, Vizcaychipi MP, Peacock S, Cinelli M, McLachlan L, Andreotti F, et al. Early risk assessment for COVID-19 patients from emergency department data using machine learning. *Sci Rep.* 2021;11:4200.
19. Hu CA, Chen CM, Fang YC, Liang SJ, Wang HC, Fang WF, et al. Using a machine learning approach to predict mortality in critically ill influenza patients: a cross-sectional retrospective multicentre study in Taiwan. *BMJ Open.* 2020;10:e033898.
20. Kim H, Jung YI, Sung M, Lee JY, Yoon JY, Yoon JL. Reliability of the interRAI Long Term Care facilities (LTCF) and interRAI Home Care (HC). *Geriatr Gerontol Int.* 2015;15:220–8.
21. Gruneir A, Bronskill S, Bell C, Gill S, Schull M, Ma X, et al. Recent health care transitions and emergency department use by chronic long term care residents: a population-based cohort study. *J Am Med Dir Assoc.* 2012;13:202–6.
22. Liu M, Maxwell CJ, Armstrong P, Schwandt M, Moser A, McGregor MJ, et al. COVID-19 in long-term care homes in Ontario and British Columbia. *CMAJ.* 2020;192:E1540–E6.
23. Brown KASN, Vanniyasingam T, et al. Early impact of Ontario's COVID-19 vaccine rollout on long-term Care Home residents and Health Care Workers. *Science Briefs of the Ontario COVID-19. Sci Advisory Table* 2021;2(13). <https://doi.org/10.47326/ocsat.2021.02.13.1.0>.
24. Garcia-Cabrera L, Perez-Abascal N, Montero-Erasquin B, Rexach Cano L, Mateos-Nozal J, Cruz-Jentoft A. Characteristics, hospital referrals and 60-day mortality of older patients living in nursing homes with COVID-19 assessed by a liaison geriatric team during the first wave: a research article. *BMC Geriatr.* 2021;21:610.
25. Heras E, Garibaldi P, Boix M, Valero O, Castillo J, Curbelo Y, et al. COVID-19 mortality risk factors in older people in a long-term care center. *Eur Geriatr Med.* 2021;12:601–7.
26. Hippisley-Cox J, Coupland CA, Mehta N, Keogh RH, Diaz-Ordaz K, Khunti K, et al. Risk prediction of covid-19 related death and hospital admission in adults after covid-19 vaccination: national prospective cohort study. *BMJ.* 2021;374:n2244.
27. Ibrahim JE, Li Y, McKee G, Eren H, Brown C, Aitken G, et al. Characteristics of nursing homes associated with COVID-19 outbreaks and mortality among residents in Victoria, Australia. *Australas J Ageing.* 2021;40:283–92.
28. Kawada T. In-Hospital mortality risk of older patients with COVID-19 infection. *J Am Med Dir Assoc.* 2022;23:1119.
29. Aryal K, Mowbray F, Gruneir A, Griffith LE, Howard M, Jabbar A et al. Nursing home Resident Admission characteristics and potentially preventable Emergency Department transfers. *J Am Med Dir Assoc.* 2021.
30. Friedman J. Greedy function approximation: a gradient boosting machine, *annals of statistics.* Ann Stat. 2001;29:1189–232.
31. Christodoulou EMJ, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol.* 2019;110:12–22.
32. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of machine learning methods with traditional models for use of administrative claims with Electronic Medical records to predict heart failure outcomes. *JAMA Netw Open.* 2020;3:e1918962.
33. Mowbray F, Zargoush M, Jones A, de Wit K, Costa A. Predicting hospital admission for older emergency department patients: insights from machine learning. *Int J Med Inf.* 2020;140:104163.
34. Feng C, Kephart G, Juarez-Colunga E, Predicting. COVID-19 mortality risk in Toronto, Canada: a comparison of tree-based and regression-based machine learning methods. *BMC Med Res Methodol.* 2021;21:267.
35. Mowbray FI, Jones A, Schumacher C, Hirdes J, Costa AP. External validation of the detection of indicators and vulnerabilities for emergency room trips (DIVERT) scale: a retrospective cohort study. *BMC Geriatr.* 2020;20:413.
36. Armstrong JJ, Zhu M, Hirdes JP, Stolee P. K-means cluster analysis of rehabilitation service users in the Home Health Care System of Ontario: examining the heterogeneity of a complex geriatric population. *Arch Phys Med Rehabil.* 2012;93:2198–205.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.