



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Factors associated with SARS-CoV-2 test positivity in long-term care homes: A population-based cohort analysis using machine learning

Douglas S. Lee,<sup>a,b,c,d,\*</sup> Chloe X. Wang,<sup>a,e</sup> Finlay A. McAlister,<sup>f,g,h</sup> Shihao Ma,<sup>a,e,i</sup> Anna Chu,<sup>a</sup> Paula A. Rochon,<sup>a,d,j</sup> Padma Kaul,<sup>g,h</sup> Peter C. Austin,<sup>a</sup> Xuesong Wang,<sup>a</sup> Sunil V. Kalmady,<sup>g</sup> Jacob A. Udell,<sup>a,b,d,j</sup> Michael J. Schull,<sup>a,d,k</sup> Barry B. Rubin,<sup>b,l</sup> and Bo Wang,<sup>a,b,e,i,m,†</sup>

<sup>a</sup>ICES (formerly Institute for Clinical Evaluative Sciences), Toronto, Canada

<sup>b</sup>Division of Cardiology, Peter Munk Cardiac Centre, Cardiovascular Program, University Health Network, University of Toronto, Toronto, ON, Canada

<sup>c</sup>Ted Rogers Centre for Heart Research, Toronto, Canada

<sup>d</sup>Department of Medicine, University of Toronto, Toronto, Canada

<sup>e</sup>Vector Institute of Artificial Intelligence, University of Toronto, Toronto, Canada

<sup>f</sup>Alberta SPOR Unit, University of Alberta, Edmonton, Canada

<sup>g</sup>Canadian VIGOUR Centre, University of Alberta, Edmonton, Canada

<sup>h</sup>Department of Medicine, University of Alberta, Edmonton, Alberta, Canada

<sup>i</sup>Department of Computer Science, University of Toronto, Toronto, Canada

<sup>j</sup>Women's College Hospital, Toronto, Canada

<sup>k</sup>Department of Emergency Medicine, Sunnybrook Health Sciences Centre, Toronto, Canada

<sup>l</sup>Division of Vascular Surgery, Peter Munk Cardiac Centre, University Health Network, Toronto, Canada

<sup>m</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada

## Summary

**Background** SARS-Cov-2 infection rates are high among residents of long-term care (LTC) homes. We used machine learning to identify resident and community characteristics predictive of SARS-Cov-2 infection.

**Methods** We linked 26 population-based health and administrative databases to identify the population of all LTC residents tested for SARS-Cov-2 infection in Ontario, Canada. Using ensemble-based algorithms, we examined 484 factors, including individual-level demographics, healthcare use, comorbidities, functional status, and laboratory results; and community-level characteristics to identify factors predictive of infection. Analyses were performed separately for January to April (early wave 1) and May to August (late wave 1).

**Findings** Among 80,784 LTC residents, 64,757 (80.2%) were tested for SARS-Cov-2 (median age 86 (78–91) years, 30.6% male), of whom 10.2% of 33,519 and 5.2% of 31,238 tested positive in early and late wave 1, respectively. In the late phase (when restriction of visitors, closure of communal spaces, and universal masking in LTC were routine), regional-level characteristics comprised 33 of the top 50 factors associated with testing positive, while laboratory values and comorbidities were also predictive. The c-index of the final model was 0.934, and sensitivity was 0.887. In the highest versus lowest risk quartiles, the odds ratio for infection was 114.3 (95% CI 38.6–557.3). LTC-related geographic variations existed in the distribution of observed infection rates and the proportion of residents at highest risk.

**Interpretation** Machine learning informed evaluation of predicted and observed risks of SARS-CoV-2 infection at the resident and LTC levels, and may inform initiatives to improve care quality in this setting.

**Funding** Funded by a Canadian Institutes of Health Research, COVID-19 Rapid Research Funding Opportunity grant (# VR4 172736) and a Peter Munk Cardiac Centre Innovation Grant. Dr. D. Lee is the Ted Rogers Chair in Heart Function Outcomes, University Health Network, University of Toronto. Dr. Austin is supported by a Mid-Career investigator award from the Heart and Stroke Foundation. Dr. McAlister is supported by an Alberta Health Services Chair in Cardiovascular Outcomes Research. Dr. Kaul is the CIHR Sex and Gender Science Chair and the Heart & Stroke Chair in Cardiovascular Research. Dr. Rochon holds the RTO/

The Lancet Regional Health - Americas 2022;6: 100146

Published online 17 January 2022

<https://doi.org/10.1016/j.lana.2021.100146>

\*Corresponding author at: Division of Cardiology, Peter Munk Cardiac Centre, Cardiovascular Program, University Health Network, University of Toronto, Toronto, ON, Canada.

E-mail address: [dlee@ices.on.ca](mailto:dlee@ices.on.ca) (D.S. Lee).

<sup>†</sup> On behalf of the CORONA collaboration.

ERO Chair in Geriatric Medicine from the University of Toronto. Dr. B. Wang holds a CIFAR AI chair at the Vector Institute.

**Copyright** © 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

**Keywords:** SARS-CoV-2; Respiratory infection; Long-term care; Elderly; Population health; Machine Learning; Artificial intelligence; Quality of care

### Research in context

#### *Evidence before this study*

The early waves of the SARS-CoV-2 pandemic saw residents of Ontario's long-term care homes experience high rates of coronavirus infection. With their older age and prevalence of comorbidities when compared with non-long-term care residents, mortality rates are also high once infected. Thus, understanding the factors that contribute to SARS-CoV-2 infection is important for proactively mitigating the risk of infection and outbreaks in these homes. Studies in the USA, England and Canada have focused on the characteristics of long-term care homes, finding that a greater number of beds, lower quality ratings, for-profit status and higher resident-to-staff ratios are associated with infection among residents. However, the role of resident characteristics and features of the community or region surrounding the long-term care home have been less well studied. The former is important for understanding why some residents get infected and others do not, while the latter is important because one study reported 23.4% of the variation in risk of infection among nursing home residents could be explained by the county in which homes are located.

#### *Added value of this study*

We examined 484 individual, long-term care home and community-level characteristics to identify factors predictive of SARS-CoV-2 test positivity during first wave of the pandemic in residents of long-term care homes in Ontario, Canada. We found that between May and August 2020, when testing and infection control policies and practices in long-term care homes were more standard and routine, 33 of the top 50 factors associated with residents testing positive were community-level characteristics including community population size and density, employment rates and immigrant population. Long-term care home features (e.g., number of beds) and individual characteristics (e.g., frailty, laboratory values) were also associated with increased infection risk. Our final model had high discrimination for predicting SARS-CoV-2 test positivity (c-index 0.934) and sensitivity (0.887) with the odds ratio for infection being 114.3 (95% CI 38.6–557.3) in the highest versus lowest risk quartiles.

### *Implications of all the available evidence*

Our examination of individual, facility and community-level factors associated with SARS-CoV-2 test positivity among long-term care home residents support prior studies' findings of a role of community characteristics in infection risk. Individual characteristics such as functional status and comorbidities are also important in explaining infection risk. Together, these factors may inform initiatives to identify high-risk facilities and communities for targeting of infection control and vaccination messages to reduce outbreaks in this setting and their surroundings.

### Introduction

The SARS-CoV-2 pandemic has had large impacts on residents of long-term care homes, who experience high mortality rates once infected with the coronavirus.<sup>1,2</sup> Consequently, understanding the upstream factors that lead to SARS-CoV-2 test positivity is very important to maintain health in long-term care homes for future SARS-CoV-2 waves or even epidemics of other respiratory infections.

To date, many studies have focused on the structural or administrative characteristics of the long-term care home itself, identifying larger home size,<sup>3</sup> lower quality ratings,<sup>4</sup> for-profit status (vs. public),<sup>5</sup> and higher resident-to-staff ratios<sup>6</sup> as being predictors of SARS-CoV-2 infection. However, few studies have examined the contributing features of the residents therein, or the effects of the community characteristics surrounding the long-term care home, and whether these factors influence infection risk. The former is important because patients' health may explain, in part, why some individuals become infected while others may not. The latter is important because prior studies have suggested that approximately 23% of the variation in SARS-CoV-2 infection rates is attributable to the county where long-term care homes are located.<sup>7</sup> However, the specific community characteristics associated with higher risk in long-term care homes have not been elucidated.<sup>8</sup>

In this study, we utilized machine learning methods to identify individual and community-level predictors of SARS-CoV-2 test positivity in residents of long-term care homes. We also compared geographic and temporal variations in observed versus predicted rates of SARS-CoV-2 test positivity in the population.

## Methods

### Data sources

We linked 26 population-based health and related administrative databases to create our study cohort, including the Ontario Laboratories Information System (OLIS) which contains test data from SARS-CoV-2 viral RNA reverse transcription (RT)-PCR testing conducted among the population of Ontario (**Appendix 1**). The Registered Persons Database was used to identify all individuals living in Ontario, alive and eligible for the Ontario Health Insurance Plan (OHIP), the province's universal health insurance available to almost all permanent residents of Ontario. Residents of a long-term care home were identified using the Ontario Drug Benefit database and Continuing Care Reporting System (CCRS). Data from CCRS' interRAI assessments were used to obtain characteristics of long-term care home residents. Immigrant status and related information were obtained from the Immigration, Refugees and Citizenship Canada Permanent Resident database. The Canadian Institute for Health Information Discharge Abstract Database (CIHI DAD) was used to obtain information on all hospitalizations and related medical diagnoses for each individual in the study cohort, using the International Classification of Diseases 10<sup>th</sup> Canadian Edition (ICD-10-CA) coding system. Where available, chronic diseases were identified using validated, disease-specific provincial databases at ICES (**Appendix 2**). We also utilized select laboratory test results based on availability and prior research (such as complete blood count components, iron and lipid measures; **Appendix 3**) conducted at any time between 2015 and 2019 and available from OLIS. Finally, regional population characteristics were obtained at the dissemination area level using data from the 2016 Canadian Census. In Canada, dissemination areas are small, relatively stable geographic areas with population sizes of 400 to 700 people bounded by the road network. Regional characteristics included information on age and sex distributions, English and French language ability, education, ethnic population, housing, and employment status.

### Study cohort

We examined all residents of long-term care homes in Ontario, Canada, who underwent testing for SARS-CoV-2 during wave 1 of the pandemic, studying 2 four-month blocks from January 1, 2020 to April 30, 2020 (early wave 1) and May 1, 2020 to August 31, 2020 (late wave 1). Our primary analysis was on late wave 1, and secondary

analysis was on early wave 1, because the beginning of the pandemic, during early wave 1, was a period of variable infection control guidelines and restricted testing policies, and staff were permitted to work at multiple facilities simultaneously. During late wave 1, SARS-CoV-2 testing of long-term care residents, broader testing of healthcare workers and caregivers, and guidelines regarding personal protective equipment to reduce transmission had become established routinely. We excluded individuals <18 years of age, non-Ontario residents and those who were not eligible for Ontario's universal health insurance plan during 2019. If an individual had repeat tests, the first positive test was considered the index test; and if no test was positive, the first test was considered the index test. This study was performed under Section 45 of Ontario's Personal Health Information and Privacy Act (PHIPA) and did not require approval by a Research Ethics Board or individual consent to be obtained. Therefore, all residents of long-term care in the province were included without participant bias.

### Covariates

In total, we examined 484 potential predictors in our predictive models, including patient demographic characteristics ( $n=27$ ), regional community features ( $n=222$ ), long-term care-related factors ( $n=2$ ), comorbidities, frailty or prior medical history-related ( $n=131$ ), interRAI functional status measures ( $n=43$ ), or laboratory tests ( $n=59$ ) as shown in **Appendix 1**. Potential covariates included individual-level demographic characteristics (e.g., age, sex, education, resident's English language ability, neighbourhood income quintile and community size of residence, likely ethnicity by surname, immigrant status), recent health care use (e.g., frequency of hospital, emergency department, and physician visits in the prior year), acute care hospital-identified comorbidities (e.g., prior organ transplants, liver disease), scores (e.g., Hospital Frailty Risk Score, Charlson score), chronic diseases (e.g., hypertension, diabetes, heart failure, cancer, COPD, asthma, need for home oxygen, chronic kidney disease, atrial fibrillation, peripheral vascular disease, rheumatoid arthritis, inflammatory bowel disease, HIV, dementia), and prior history of infectious respiratory conditions (e.g., pneumonia or influenza, SARS infection, 2009 H1N1 infection, respiratory tuberculosis, invasive pneumococcal disease or other acute respiratory infections), with diagnostic codes shown in **Appendix 2**.

We also examined outcome scales and clinical assessment protocols implemented from the interRAI Resident Assessment Instrument (RAI-MDS 2.0), which is used to assess the clinical status of long-term care residents and identify areas for potential intervention,<sup>9</sup> and laboratory test results in the province-wide Ontario Laboratory Information System (OLIS) database (see **Appendix 3** for a list of all laboratory tests considered). Since prior reports

identified the importance of neighbourhood on the likelihood of SARS-CoV-2 infection, we included regional socio-demographic characteristics at the dissemination area level, with 20,160 dissemination areas in the province. Potential predictors were included in our models only if they were present prior to the date of the index SARS-CoV-2 test.

### Analysis

Our analyses were performed in the Ontario Health Data Platform environment at ICES, a secure private cloud-based platform, where encoded personal health information linked to ICES' population datasets described above can be analyzed using machine learning. We trained our data using the XGBOOST algorithm to predict SARS-CoV-2 test positivity, formulated as a binary classification problem, in the long-term care population who underwent testing. The XGBOOST model is a decision-tree-based ensemble method that uses sequential decision trees and gradient descent to optimize predictive performance.<sup>10</sup> In a decision tree, the predictor space is partitioned into segments and each observation (i.e., individual) is classified into a segment. In each iteration of boosting, a new tree is added to the existing sequence of decision trees to further correct the errors of the previous trees.

For early and late wave 1 separately, we split the data into training (80%) and testing (20%) sets. We then randomly subsampled the negative cases in the training set to match with the positive cases in a ratio of 1:1, to construct a balanced training set on which the XGBOOST model was trained, with optimal hyperparameters identified from 5-fold cross validation.<sup>11</sup> This process of random subsampling and training was repeated 100 times to produce 100 trained models. The final model, consisting of the averaged probability outputs from the 100 trained models, was evaluated on the testing set. Where continuous data were missing, we imputed using the cohort mean after standardization (age/sex stratified mean for laboratory values). Since we used one-hot encoding (i.e., dummy variables) to expand categorical variables into  $n$  levels for analysis, missing data were automatically encoded as 0. In sensitivity analyses, we also modeled test positivity among the entire long-term care population, grouping untested and never testing positive individuals together. We reported the c-index, sensitivity and specificity with 95% confidence intervals.

To identify the most important features for the optimized model, we used the SHAP interpretability method where, for each trained model, a SHAP score was computed for each feature to explain its contribution to individual prediction.<sup>12</sup> The final SHAP importance score for each feature was then calculated as the mean absolute SHAP value from the 100 models. Spearman's rank correlation analysis was performed on the predictors and SHAP scores to identify the

directionality of association. Using the optimized model, we divided the cohort into risk quartiles, and determined the odds ratio for SARS-CoV-2 test positive versus negative status, using the lowest risk quartile as the reference category. We used the  $\chi^2$  test or Fisher's exact test to compare categorical variables.

We examined the predicted versus observed rates of ever testing SARS-CoV-2 positive in two ways. First, for each long-term care home in Ontario, we constructed a plot of observed test positivity rates during late wave 1 and contrasted it with the predicted prevalence of those who were at highest risk for SARS-CoV-2 infection (risk quartile 4) to determine if the model provided additive information beyond simple infection rates. Second, we plotted the predicted versus observed SARS-CoV-2 positivity rates on a two-dimensional grid, and using the average disease prevalence in long-term care homes of ~20%, we divided homes into four categories: (i) higher than average predicted risk and higher than average observed rates ['red quadrant'], (ii) higher than average predicted risk but lower than average observed rates ['yellow quadrant'], (iii) lower than average predicted risk and higher than average observed rates ['white quadrant'], and (iv) lower than average predicted risk and lower than average observed rates ['green quadrant']. Analyses were conducted using Python version 3.7.4 (Scotts Valley, USA), R version 3.6.2 (Vienna, Austria), and SAS version 9.4 (Cary, North Carolina). Maps were constructed using ArcGIS Desktop version 10.7 by ESRI.

### Role of the funding source

The funders of this study had no role in the study design, data collection, data analysis, interpretation or writing of this manuscript.

## Results

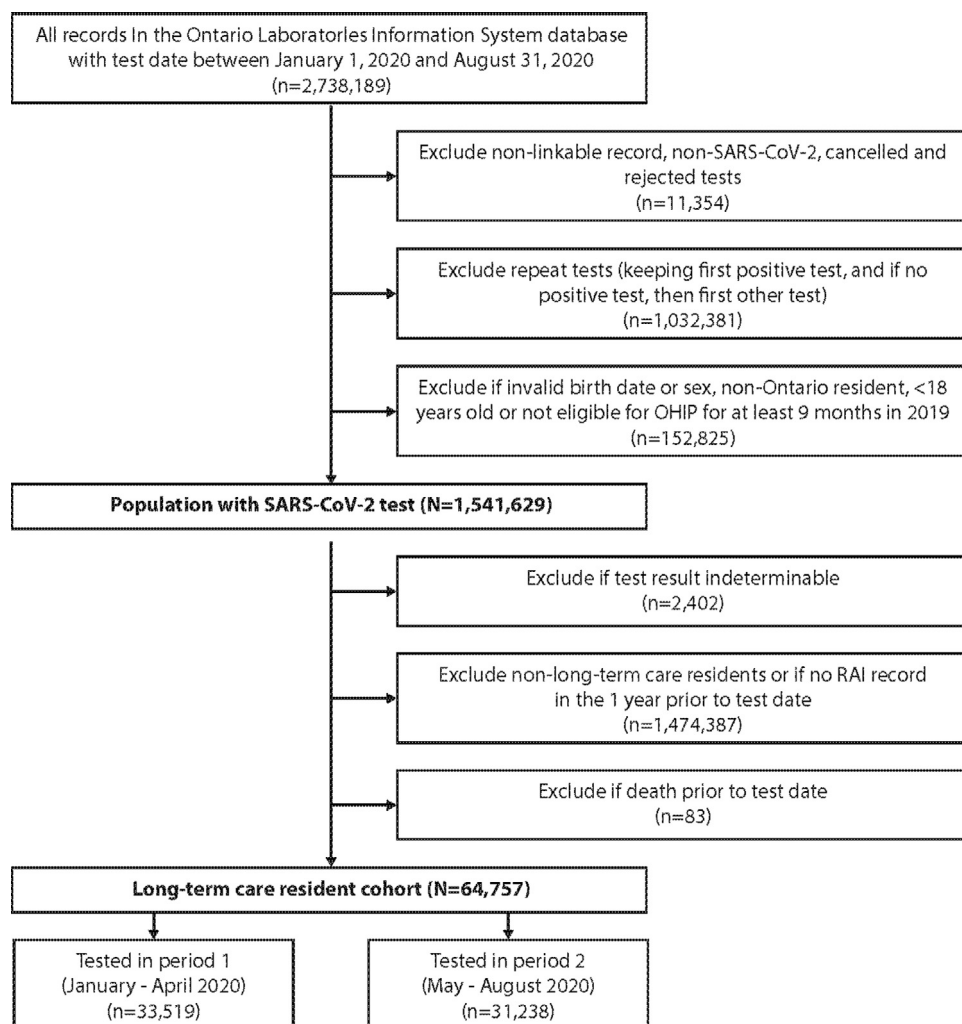
### Study cohort

Among 80,784 residents of long-term care homes, 64,757 residents underwent testing for the SARS-CoV-2 virus during the study period. A study flow diagram is shown in [Figure 1](#). Of the 33,519 (51.8%) residents tested in early wave 1 (Jan to Apr 2020), 3409 (10.2%) tested positive; and of the 31,238 (48.2%) residents tested in late wave 1 (May to Aug 2020), 1627 (5.2%) tested positive. Abridged baseline cohort characteristics for all long-term care residents tested are shown in [Table 1](#) and an expanded list, including characteristics of those who were never tested, is shown in [Appendix 4](#). Most long-term care residents (80.2%) in Ontario were tested for SARS-CoV-2.

### Characteristics

A summary of the top 50 features associated with a resident testing positive in late wave 1 is shown in [Table 2](#),





**Figure 1.** Study flow diagram.

and associations in early wave 1 are shown in **Appendix 5**. Almost all features had a  $p$ -value  $< 0.001$  for Spearman's correlation coefficient, indicating the direction of the association. A detailed explanation of these features is shown in **Appendix 6**. The features associated with test positivity in the late phase included demographic characteristics ( $n=3$ ), comorbid conditions ( $n=4$ ), functional characteristics ( $n=2$ ), regional features ( $n=33$ ), nursing home ( $n=1$ ) and laboratory test features ( $n=7$ ). Regional features were related to an individual's residence at the start of the study period and were distinct from the location of the long-term care home itself in 26,710 (41.2%) of all individuals.

The final model had high discrimination for predicting positivity for SARS-CoV-2 (c-index 0.934) in those who underwent testing. Similarly, the model had high discrimination for predicting positivity in all residents (including those who underwent testing and those who did not) (c-index 0.934). Sensitivities were 0.887 and

0.865, respectively, and specificities were 0.869 and 0.862, respectively (see **Table 3** for details). Performance of this model in early wave 1 was similar (c-indices 0.877 and 0.913, sensitivities 0.794 and 0.857, and specificities 0.815 and 0.831, **Appendix 7**).

#### Stratification of risk quartiles

**Table 4** shows odds ratios for test positivity by risk quartile in late wave 1. Compared to the lowest risk quartile, those in the highest risk quartile had over 110-fold risk when only tested persons were considered. In the entire long-term care population, including untested individuals, those in the highest risk quartile had over 150-fold risk. Similar findings were observed in early wave 1 irrespective of whether the non-positive cohort were those who tested negative or test-negatives combined with untested individuals (**Appendix 8**).

Characteristic	Units or range	n (%) or median (25th, 75th %iles)
<b>Demographic characteristics</b>		
Age	Years	86 (78, 91)
Male	%	9,555 (31%)
Education secondary school or less	Yes	12,412 (40%)
	No	5194 (17%)
	Unknown	13,632 (44%)
<b>Community characteristics</b>		
Community size*	1,500,000+	7291 (23%)
	500,000 to 1,499,999	7005 (22%)
	100,000 to 499,999	7966 (26%)
	10,000 to 99,999	4721 (15%)
	≤ 10,000	4144 (13%)
Census subdivision population density	Density per km <sup>2</sup>	480 (146, 1428)
Postal code population size	Count	129 (88, 215)
Dissemination area population size	Count	664 (523, 1017)
Dissemination area population density	Density per km <sup>2</sup>	1986 (710, 3718)
Regional sex distribution	Male population size	45.7 (42.5, 48.8)
	Female population size	54.1 (51.2, 57.5)
Age-stratified distribution, men	% who are 0 to 44 years	21.8 (16.6, 27.0)
	% who are 45 to 64 years	11.9 (9.8, 14.1)
	% who are ≥ 65 years	11.3 (7.7, 15.3)
Age-stratified distribution, women	% who are 0 to 44 years	21.7 (16.5, 26.3)
	% who are 45 to 64 years	13.2 (11.0, 15.3)
	% who are ≥ 65 years	17.4 (9.8, 27.6)
Regional family units, % of population	% Never married	23.1 (18.5, 28.2)
	% Married or common-law	50.5 (42.8, 59.7)
	% Separated or divorced	9.1 (7.1, 12.1)
	% Widowed	13.5 (5.9, 21.7)
Ontario marginalization index	Ethnic concentration score	-0.49 (-0.81, 0.19)
	Deprivation factor score	-0.21 (-0.68, 0.52)
Visible minorities	%	9.9 (2.9, 24.9)
Occupied private dwellings that are apartments in buildings and other attached dwellings	%	54.2 (20.2, 97.2)
Post-secondary education	% with none	81.7 (66.1, 99.3)
Languages spoken	English but not French	90.6 (84.1, 93.3)
	English and French	7.3 (5.0, 12.4)
	French but not English	0.0 (0.0, 0.0)
	Neither English or French	0.7 (0.0, 2.0)
Non-immigrants	%	80.8 (66.4, 91.1)
% of immigrants who immigrated in specified years	Prior to 1981	7.4 (4.4, 12.2)
	1981 to 1990	2.1 (0.0, 4.0)
	1991 to 2000	2.4 (0.0, 6.2)
	2001 to 2010	2.5 (0.0, 6.4)
	2011 to 2016	0.0 (0.0, 3.0)
Non-permanent residents	%	0.0 (0.0, 1.7)
% Employed in DA by sex	Men	60.0 (51.7, 68.8)
	Women	53.4 (44.6, 62.0)
<b>Long-term care home characteristics</b>		
Bed size	# of beds	152 (104, 197)
Time in residence	Days	608 (262, 1211)
Payment source <sup>†</sup>	Government	27,531 (88%)
	Private	393 (1%)
	Self	22,949 (73%)

(continued)

Table 1 (Continued)

Characteristic	Units or range	n (%) or median (25th, 75th %iles)
	Unknown	1,823 (6%)
<b>Comorbidities and health status</b>		
COPD duration, if present	Years	9.7 (4.7, 17.5)
Coronary revascularization, if received	Time since last procedure	11.6 (7.1, 16.5)
Dementia duration, if present	Years	5.0 (2.8, 8.2)
ED visits with a respiratory diagnosis	# of visits in prior year	0 (0, 0)
ED visits any reason	# of visits in prior year	0 (0, 1)
ED visits with transfer to residential care	# of visits in prior year	0 (0, 0)
Frailty	Hospital Frailty Risk Score	-1 (-1, 5.9)
Hypertension duration, if present	Years	19.9 (13.6, 26.4)
Heart failure duration, if present	Years	5.7 (2.7, 11.3)
Hospitalizations for respiratory infection	# of hospitalizations since 2000	0 (0, 0)
Primary care physician visits	# in prior year	12 (11, 13)
<b>Functional (InterRAI) status</b>		
ADL scale – long form	Range: 0–28 Higher score: greater impairment of self-sufficiency	19 (14, 22)
ADL self-performance hierarchy	Range: 0=independent to 6=total dependence	4 (3, 5)
Aggressive behavior scale	Range: 0=low to 12=high	0 (0, 2)
CHESS (Changes in Health, End-stage disease, Symptoms & Signs)	Range: 0=no instability to 5=highest instability	0 (0, 1)
Cognitive performance scale	Range: 0=lowest risk to 6=highest risk	3 (2, 4)
Dehydration risk	Present	2661 (9%)
Depression rating scale	Range: 0–14	1 (0, 3)
Index of social engagement	Range: 0=low level of social engagement, 6=high level	3 (2, 4)
Pressure ulcer risk scale	Range: 0=lowest risk to 8=highest risk	2 (1, 3)
Stage 2+ pressure ulcer	Protocol initiated=Yes	1576 (5%)
Undernutrition	Medium or high risk	5195 (17%)
Urinary incontinence	Present	21,298 (68%)
<b>Laboratory tests*</b>		
Bilirubin	μmol/L	8 (5, 11)
Calcium, total	mmol/L	2.30 (2.22, 2.39)
GFR	mL/min/1.73 m <sup>2</sup>	65 (49, 80)
Hemoglobin	g/L	122 (111, 132)
High density lipoprotein	mmol/L	1.22 (0.99, 1.52)
Iron saturation	Percent	0.22 (0.16, 0.30)
Lymphocyte count	×10 <sup>9</sup> /L	1.6 (1.2, 2.1)
Platelet count	×10 <sup>9</sup> /L	231 (188, 281)
Serum albumin	g/L	38 (35, 41)
Serum potassium level	mmol/L	4.3 (4.0, 4.6)
Magnesium	mmol/L	0.84 (0.77, 0.90)
Thyroid stimulating hormone	mIU/L	2.03 (1.31, 3.04)
Thyroxine (Free T4)	pmol/L	14.0 (12.0, 16.0)
Total cholesterol	mmol/L	4.01 (3.29, 4.87)
Total:HDL cholesterol ratio		3.20 (2.58, 4.00)
Total iron binding capacity	μmol/L	51.8 (44.2, 59.0)
Urate	μmol/L	324 (265, 393)

Table 1: Baseline characteristics of study cohort (late phase, N = 31,238).

ADL = activities of daily living, COPD = chronic obstructive pulmonary disease, ED = emergency department, GFR = glomerular filtration rate, HDL = high density lipoprotein

\* Community size unknown for 111 individuals (0.4%). Missing laboratory data varied from 0.2% for eGFR to 92.5% for ionized calcium.

† May have more than one source of funding.



#	Characteristic and description	Type of feature	Spearman's rank correlation	Relationship with SARS-CoV-2 infection
1	Community size	Community	-0.8142	Larger community size correlated with higher infection risk
2	Census subdivision population density	Community	0.8917	Higher population density correlated with higher infection risk
3	Number of long-term care beds at the home	LTC-related	0.1993	More beds correlated with higher infection risk
4	Percent never married†	Community	0.8335	Higher % never married correlated with higher infection risk
5	Percent visible minority population†	Community	0.8962	Higher % visible minority correlated with higher infection risk
6	Female employment rate†	Community	0.2031	Higher % women employed correlated with higher infection risk
7	Percent male 45 to 64 years†	Community	0.8170	Higher % men in age group correlated with higher infection risk
8	Hospital frailty risk score	Comorbidity	0.7948	Increased frailty correlated with higher infection risk
9	Total female population size†	Community	-0.8909	Greater female population correlated with lower infection risk
10	Percent immigrated 1981 to 1990†	Community	0.8960	Higher % immigrant from 1981 to 1990 correlated with higher infection risk
11	Percent non-permanent residents†	Community	0.7115	Higher % non-permanent residents correlated with higher infection risk
12	Dissemination area population size†	Community	0.2995	Larger population size in DA correlated with higher infection risk
13	Percent who speak French and not English†	Community	0.0160	Higher % French-only speakers correlated with higher infection risk
14	Male employment rate†	Community	0.7607	Higher employment rate correlated with higher infection risk
15	Percent immigrated prior to 1981†	Community	0.4582	Higher % immigrant prior to 1981 correlated with higher infection risk
16	Postal code population size	Community	0.5528	Larger population in postal code region correlated with higher infection risk
17	ONMARG ethnic concentration score†	Community	-0.3618	Lower ethnic concentration correlated with higher infection risk
18	ONMARG deprivation factor score†	Community	0.6639	Greater community deprivation correlated with higher infection risk
19	Dissemination area population density†	Community	0.7251	Higher population density in DA correlated with higher infection risk
20	Percent who speak English and French†	Community	0.7156	Higher % speakers of English and French correlated with higher infection risk
21	Percent of occupied private dwellings that are apartments in buildings and other attached dwellings†	Community	-0.7036	Higher % dwellings that are apartments in buildings and other attached dwellings correlated with lower infection risk
22	Total male population size†	Community	0.7493	Higher % male population correlated with higher infection risk
23	Surname-based Chinese ethnicity	Demographic	-0.3654	Chinese ethnicity by surname correlated with lower infection risk
24	Secondary school education only	Demographic	-0.8454	Less than secondary education correlated with higher infection risk
25	Percent female 0 to 44 years†	Community	-0.8770	Higher % women in age group correlated with lower infection risk
26	Percent who speak English and not French†	Community	-0.0173	Higher % speakers of English correlated with lower infection risk
27	Percent who speak neither English or French†	Community	-0.6941	Higher % non-English, non-French speakers correlated with lower infection risk

(continued)

Table 2 (Continued)

#	Characteristic and description	Type of feature	Spearman's rank correlation	Relationship with SARS-CoV-2 infection
28	Duration of hypertension	Comorbidity	-0.7811	Longer duration of hypertension correlated with lower infection risk
29	Percent female 45 to 64 years <sup>†</sup>	Community	0.8562	Higher % women in age group correlated with higher infection risk
30	Percent without post-secondary education <sup>†</sup>	Community	0.6004	Higher % without post-secondary education correlated with higher infection risk
31	Number of days in long-term care	LTC-related	-0.9633	Longer days in long-term care correlated with lower infection risk
32	Magnesium	Laboratory	0.1939	Higher magnesium level correlated with higher infection risk
33	Percent immigrated between 2001 and 2010 <sup>†</sup>	Community	0.5002	Higher % immigrant from 2001 to 2010 correlated with higher infection risk
34	Percent of population male 65+ years <sup>†</sup>	Community	-0.1600	Higher % men in age group correlated with lower infection risk
35	Thyroxine (Free T4)	Laboratory	-0.9267	Higher thyroxine correlated with lower infection risk
36	Bilirubin	Laboratory	0.7387	Higher bilirubin correlated with lower infection risk
37	ED visits resulting in transfer to residential care	Comorbidity	0.5886	Higher # of ED visits in prior year correlated with higher infection risk
38	Total calcium	Laboratory	-0.9539	Lower calcium correlated with higher infection risk
39	Depression Rating Scale	Functional	-0.5050	Lower depression score correlated with higher infection risk
40	Percent immigrated 2011 to 2016 <sup>†</sup>	Community	-0.8057	Higher % immigrant from 2011 to 2016 correlated with higher infection risk
41	Percent widowed <sup>†</sup>	Community	-0.7014	Higher % widowed correlated with lower infection risk
42	Thyroid stimulating hormone	Laboratory	0.4098	Higher TSH correlated with higher infection risk
43	eGFR	Laboratory	0.5391	Higher eGFR correlated with higher infection risk
44	Percent separated or divorced <sup>†</sup>	Community	0.7892	Higher % separated or divorced correlated with higher infection risk
45	Percent non-immigrants <sup>†</sup>	Community	-0.3616	Higher % non-immigrants correlated with lower infection risk
46	Number of prior family doctor visits	Comorbidity	0.5159	More visits in year prior to index date correlated with higher infection risk
47	Pressure Ulcer Risk Scale (PURS)	Functional	0.8601	Higher pressure ulcer risk correlated with higher infection risk
48	Duration of dementia	Comorbidity	0.4323	Longer duration of dementia correlated with higher infection risk
49	Number of ED visits in prior year	Comorbidity	0.7707	Higher # of ED visits in prior year correlated with higher infection risk
50	Percent married or common-law <sup>†</sup>	Community	0.0868	Higher % married or common law correlated with higher infection risk

Table 2: Top 50 features associated with SARS-CoV-2 infection in late wave 1\*.

ED = emergency department, LTC = Long-term care, ONMARG = Ontario Marginalization Index

\* All features have a p-value &lt; 0.001 except percent of population with knowledge of French and not English (p=0.011)

<sup>†</sup> Based on the 2016 Canadian Census and at the dissemination area (DA) level. DA's are small, relatively stable geographic areas with populations sizes of 400 to 700 people bounded by the road network and determined from Statistics Canada's Census of Population Program.

Population	C-index(95% CI)	Sensitivity(95% CI)	Specificity(95% CI)
Tested residents only	0.934 (0.915, 0.951)	0.887 (0.854, 0.915)	0.869 (0.857, 0.878)
Tested + untested residents*	0.934 (0.918, 0.949)	0.865 (0.831, 0.908)	0.862 (0.854, 0.868)

Table 3: Performance of ML models for SARS-CoV-2 infection in long-term care residents in late wave 1 in Ontario

\* Untested residents were considered to be test negative; c-index represents discrimination of SARS-CoV-2 positive vs. test negative or not tested

Test Positivity – Full models	Total N	Positives,n (%)	Odds Ratio (95%CI)	p-value
Tested residents only				
Risk quartile 1 (lowest)	1483	<6	Reference	NA
Risk quartile 2	1448	5–9	2.40 (0.55, 14.39)	0.221
Risk quartile 3	1728	17 (1.0%)	4.90 (1.41, 26.15)	0.006
Risk quartile 4	1594	300 (18.8%)	114.26 (38.58, 557.27)	< 0.001
Tested + untested residents*				
Risk quartile 1 (lowest)	2174	≤6	Reference	NA
Risk quartile 2	2333	3–7	2.80 (0.50, 28.40)	0.291
Risk quartile 3	2527	20 (0.8%)	8.66 (2.10, 76.42)	< 0.001
Risk quartile 4	2416	299 (12.4%)	153.30 (41.98, 1241.07)	< 0.001

**Table 4: Positivity for SARS-CoV-2 based on risk quartiles from full model (all features + OLIS) in late wave 1**

NA = not applicable

\* Untested residents were considered to be test negative; odds ratios represent odds of being SARS-CoV-2 test positive vs. test negative or not tested

### Predicted versus observed long-term care home analysis

SARS-CoV-2 positivity rates by long-term care home during late wave 1 are shown in [Figure 2](#). Of 557 long-term care homes with available data, the 144 (25.9%) with at least one positive case of SARS-CoV-2 were dispersed throughout the province, but homes with the highest rates were observed in selected foci. [Figure 3](#) shows the distribution of long-term care homes by proportion of residents in the highest risk quartile in late wave 1, which shows greater dispersion of high-risk patients in a greater number of dissemination areas than rates of positivity shown in [Figure 2](#).

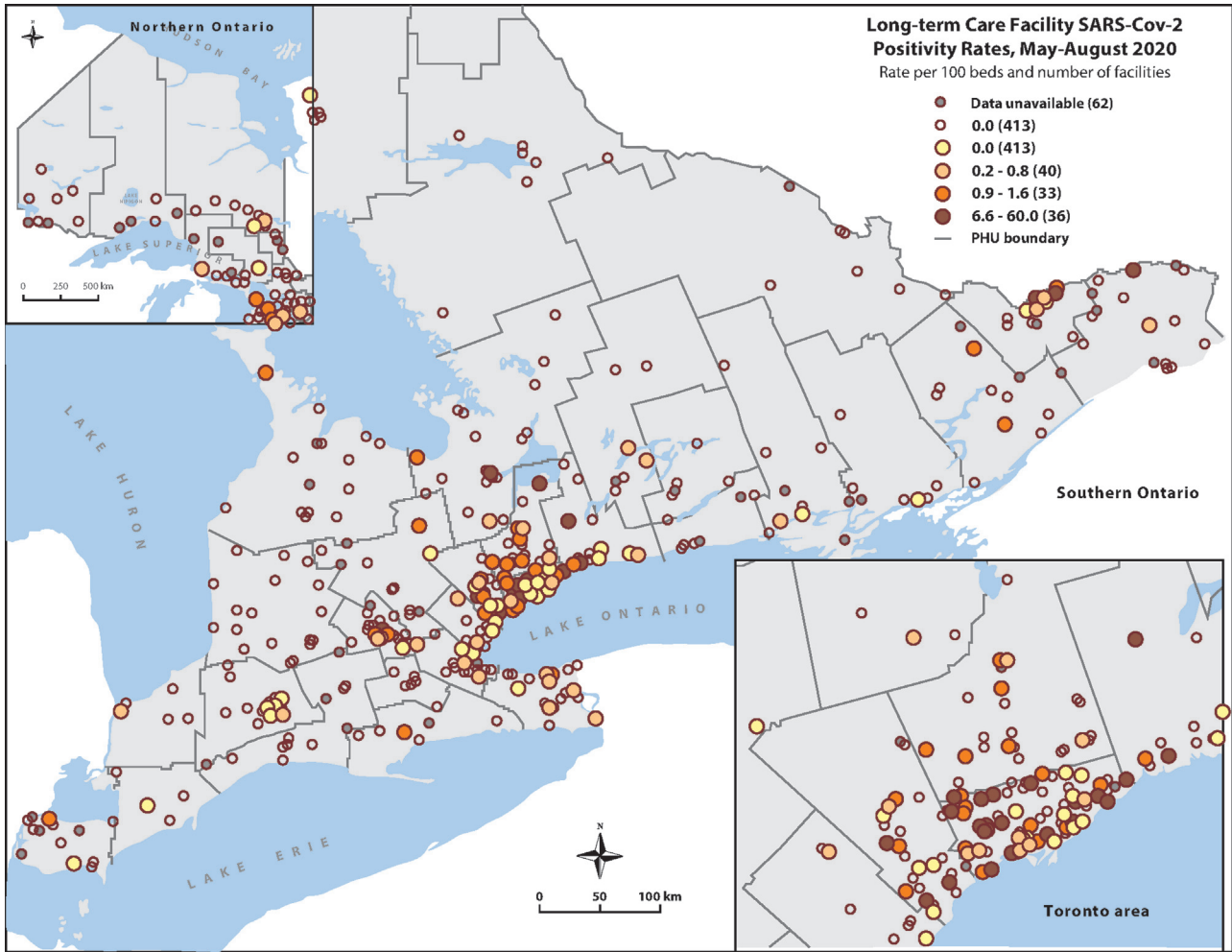
Plots of percent positivity at the long-term care home level versus the proportion of residents in the two highest quartiles during early and late wave 1 are shown in [Figure 4a](#) and [4b](#), respectively. Using as a landmark the mean SARS-CoV-2 test positivity rate of 19% at long-term care homes with at least one case during early wave 1, (dashed vertical line in [Figure 4a](#) and [b](#)), many long-term care homes had over 50% of residents in the highest predicted risk groups (quartiles 3 and 4). However, not all had an observed positivity rate that was higher than average. Long-term care homes in the yellow quadrant had an observed positivity rate that was lower than the mean yet over 50% of their residents were in risk quartiles 3 or 4. During late wave 1, despite long-term care homes having a range of predicted risks from 0 to 100%, the long-term care percent positivity rates improved and many were shifted left, clustering towards rates < 10%. Some long-term care homes had persistently high positivity rates (red quadrant, [Figure 4b](#)), while others showed high rates despite lower proportions of

patients in the highest two risk quartiles (white quadrant, [Figure 4b](#)).

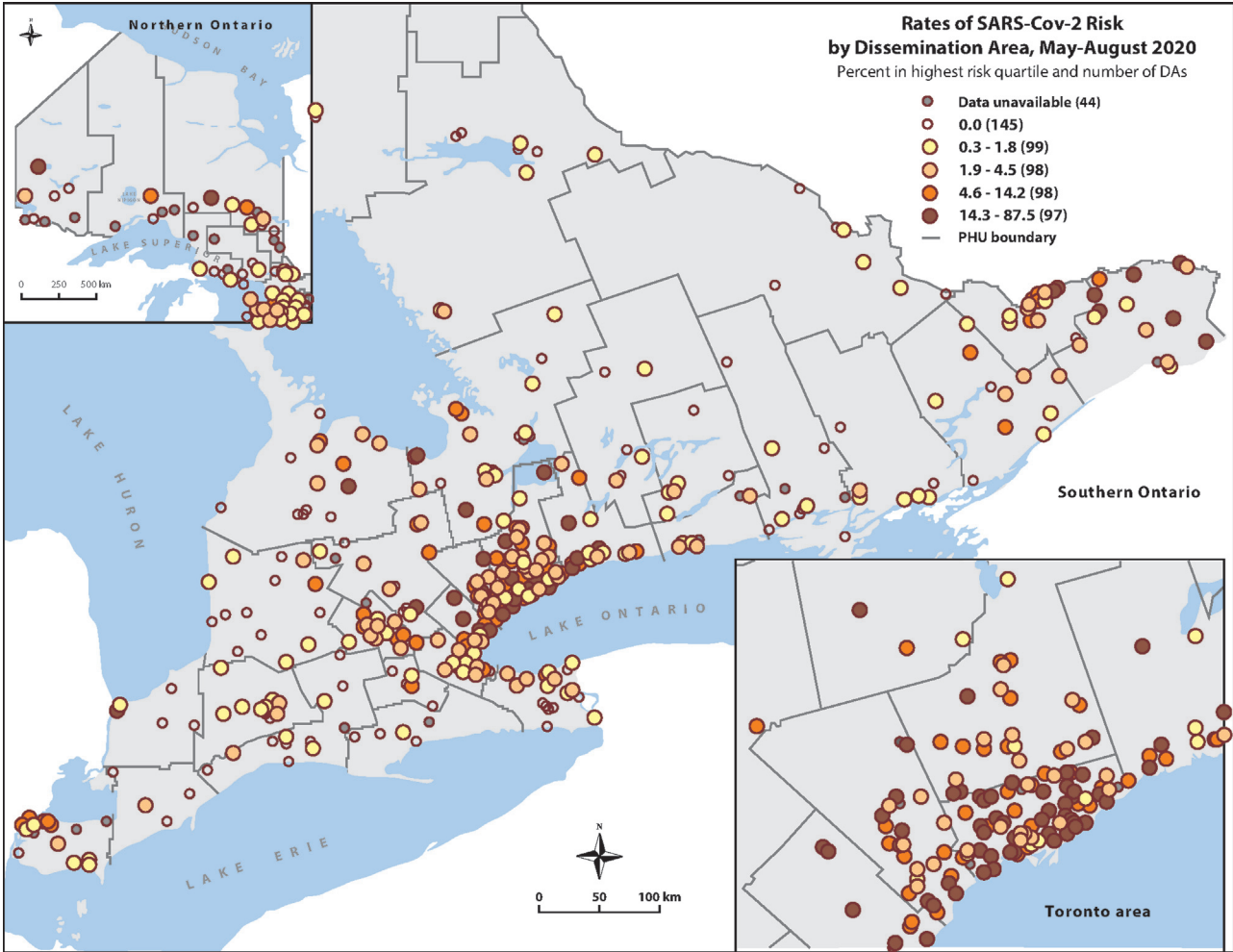
### Discussion

In this population-based analysis of long-term care residents, we found that regional characteristics were major predictors of SARS-Cov-2 infection (33 of the top 50 factors associated with infection rates). Clinical factors associated with infection included general measures of overall health service use, such as primary care and emergency department visits and frailty, but also included specific conditions, such as dementia. Model discrimination was robust, irrespective of whether controls included all non-positive patients or were limited to test negative residents who underwent testing. The model stratified the risk of SARS-CoV-2 test positivity by over 100-fold in the highest risk quartile, with an absolute risk of 0.3% or less in the lowest risk quartile. Interestingly, during the two phases we studied (January to April and May to August), we found temporal differences in the predicted versus observed rates of SARS-CoV-2 test positivity. Specifically, while high rates of SARS-CoV-2 test positivity occurred in some long-term care homes, high-risk residents were distributed more broadly geographically, indicating that structural (e.g., multiple-occupancy rooms, shared bathrooms) and process indicators (e.g., infection control practices) may have had a contributory role in the observed discordance between predicted risks and observed test positive rates.

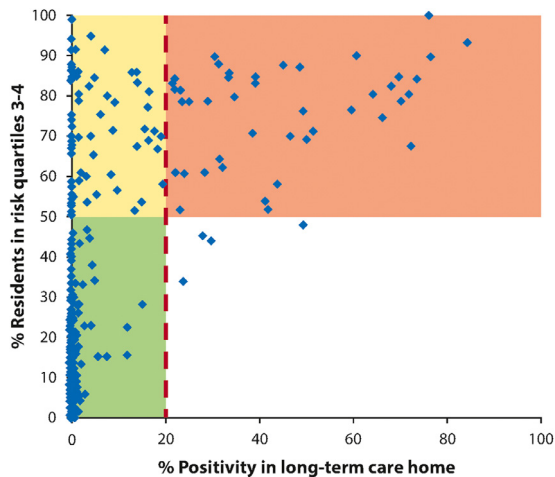
Asymptomatic transmission of SARS-CoV-2 underscores the importance of a broad approach to the consideration of potential factors that may increase transmissibility or susceptibility to infection.<sup>13</sup> A prior



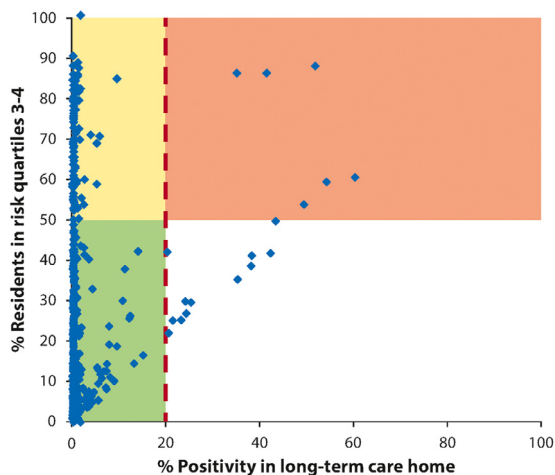
**Figure 2.** SARS-CoV-2 positivity rates by long-term care home during late wave 1.



**Figure 3.** Proportion of residents in the highest risk quartile by long-term care home in late wave 1.



**Figure 4a.** Long-term care home positivity rates versus proportion of residents in the two highest quartiles during early wave 1.



**Figure 4b.** Long-term care home positivity rates versus proportion of residents in the two highest quartiles during late wave 1.

study that examined 12,576 institutions in the US, found that infection rates at long-term care homes were associated with local county counts of Covid-19 infection, deaths, population size, and sociodemographic factors.<sup>8</sup> Sun et al. conducted an analysis of long-term care homes from four US states using machine learning, and found that some of the strongest predictors of Covid-19 infection were institution size, population density of the county, and community rates of infection in the long-term care home's county.<sup>14</sup> However, these studies did not explore details of the community characteristics which predisposed long-term care homes to a higher risk. Our study expands upon prior studies by delving further into specific features of the community, and concomitantly examining numerous patient characteristics. For example, higher employment rates, proportionately more

individuals of working age, lower proportion of those with post-secondary or university education, and greater marginalization may lead to higher community rates, particularly since these individuals may be employed in jobs with high risk of SARS-CoV-2 exposure in the workplace.

Several prior studies examining the risk of SARS-CoV-2 infection in long-term care homes, while informative, have not examined patient risk factors for infection at a population level. In an academic long-term chronic care facility in Massachusetts, predictors of Covid-19 infection included bowel incontinence and staff residence in a community with high burden of Covid-19, indicating that close, prolonged contact may lead to transmission from staff.<sup>15</sup> Other studies that have been population-based examined structural or institutional characteristics, but did not fully evaluate patients' clinical characteristics including functional status and laboratory test findings.<sup>16</sup> In an ecological study of long-term care homes in Connecticut, a higher concentration of Medicaid residents and those of racial/ethnic minorities overall were associated with higher rates of infection at the institutional level.<sup>4</sup> Our study expands on prior work by showing that numerous clinical, laboratory and health status-related factors additionally predict higher risk of SARS-CoV-2 infection.

Many factors associated with higher rates of infection in long-term care homes were community-related. While beyond the scope of this study, other studies have examined without clear results the role of aerosolization and viral dissemination through air,<sup>17</sup> and survival of SARS-CoV-2 on surfaces for prolonged periods including medical and non-medical surfaces (e.g., door handles and ATM machines).<sup>18,19</sup> Another possibility is that healthcare workers who staff long-term care homes may become exposed to SARS-CoV-2 when traveling in public areas including public transit. A connection between human mobility and infectious disease dynamics has been described, with risks escalating when those who reside in the region are also of working age, and need to travel to work because they are unable to work from home. These risks are further accentuated by increasing population density in the region, making physical distancing more difficult. Indeed, in an analysis of Ontario data, we also found that postal code regions with high rates of use of public transit for work-related travel exhibited high rates of SARS-CoV-2 infection in the region (unpublished data). In a study of long-term care homes in France, of 17 homes in which staff members confined themselves to the home with the residents, Covid-19 infections occurred in only one home (5.8%).<sup>20</sup> However, where staff members did not confine themselves with residents, 48% of long-term care homes had cases of Covid-19 indicating the propensity to introduce community-acquired infections into the homes.<sup>20</sup> In addition to regional characteristics, larger long-term care homes were associated with higher risk



of SARS-CoV-2 infection, which may be related, in part, to spread from the outside community at higher rates than smaller institutions. Interestingly, functional measures, such as pressure ulcer risk may reflect mobility limitations which require assistance from healthcare workers, who may have been infected, yet asymptomatic at the time of close contact.

Our population-based study has three major implications. First, community-based characteristics were major predictors of SARS-CoV-2 infection in residents of long-term care homes, such as community size, population density and employment rate. These area-level characteristics may reflect the inherent risk of SARS-CoV-2 infection and transmission in the community, underscoring the importance of preventative strategies such as reducing inter-institutional work among healthcare workers. Second, our geographical analysis showed that even though some long-term care homes had proportionately fewer high-risk patients, their % positivity rates were higher than average, indicating that structural, process or community factors (e.g., more beds or shared rooms, poor ventilation) may be contributory. Conversely, long-term care homes with more high-risk residents whilst having lower % positivity rates may demonstrate high quality infection control or preventive measures. In long-term care homes with a high proportion of high-risk residents who also experience high rates of infection, individual level factors (e.g., education, language ability, physical function) may be important considerations for assessing disease susceptibility. Finally, our temporal analysis demonstrated that while many long-term care homes with a high proportion of high-risk residents had high rates of infection in early wave 1, they had lower than average rates of infection in late wave 1. However, some long-term care homes continued to have higher than average rates of infection in the late phase. Consequently, continuous training using routinely collected data to reflect changes in long-term care home and resident characteristics may assist in identifying major future events. Further study of the differences between these institutions and their care practices may help prepare long-term care homes for future waves or future pandemics.

In addition to our use of detailed information on residents and the community in which they live, a strength of our population-based analysis was the mass screening of long-term care residents, such that 80.2% of identified individuals residing in homes were tested during the timeframe of our study. While testing in early wave 1 was primarily restricted to high-risk individuals (e.g., symptoms, contact with a positive case), by late wave 1, proactive and routine testing of priority groups including long-term care residents had been initiated in Ontario which may have contributed to the lower positivity rate in late versus early wave 1. Although we could not determine the reasons for testing, this is important because we were able to capture positive

SARS-CoV-2 tests even in those with minimal symptoms or in those who are asymptomatic.<sup>21</sup> However, there were some notable limitations. First, we did not have details on the structural characteristics of the long-term care homes such as use of shared bathrooms and measures of crowding, which prior studies have associated with SARS-CoV-2 transmission. However, we did examine residents' source of funding for their stay, either part private or complete government funding, which contributes to residents' level of privacy of living arrangements, and this was not significant. On the other hand, size of the long-term care home was an important predictor. Second, our primary analyses were restricted to residents with a known positive or negative SARS-CoV-2 test. With testing restrictions during early wave 1 and some differences in characteristics between the tested and non-tested population (e.g., older age and longer time in long-term care among non-tested), it is possible that exclusion of non-tested residents introduces collider bias into the early wave 1 results.<sup>22,23</sup> Collider bias may occur when both the risk factor (e.g. some high risk comorbidities) and outcome of interest (test positivity) are related to and conditioned upon a third related variable, i.e., SARS-CoV-2 testing. Hence, we chose to focus on late wave 1, and additionally performed sensitivity analyses grouping untested and test negative residents to examine model performance, which also performed well. Third, since not all residents may have received all laboratory tests included in this analysis, we chose to impute these values when missing, which could have impacted our results. Fourth, we did not have information about healthcare workers' positivity rates or characteristics, and they may have unknowingly been vectors mediating transmission to long-term care residents. Lastly, as we included 484 individual and community-level risk factors in our analyses, it is unknown whether similar results would be found in other jurisdictions or whether such data would be available elsewhere to enable similar analyses. However, identification of factors associated with increased infection risk provide public health policy makers insights into identifying high-risk communities for targeting of infection control and vaccination messages, and researchers elsewhere with methods and features for consideration and adaptation in their analyses. Additionally, despite these limitations, our models exhibited excellent c-indices indicating that many important factors were included.

## Conclusions

In conclusion, in addition to patient-specific factors, regional community-based factors were predictive of SARS-CoV-2 test positivity in long-term care residents. Geographic and temporal variations in the risks and observed rates of SARS-CoV-2 infection exist, and need to be explored further to understand the contributions

of structural and process measures at institutions with discordance between predicted risks and observed rates of infection.

### Contributors

All authors contributed to the concept, study design and interpretation of results. DSL, CXW, SM, AC, PCA, XW, MJS and BW contributed to data acquisition and analysis. DSL wrote the first draft of the manuscript, and all authors reviewed and revised it for important intellectual content and provided approval of the final version.

### Data sharing statement

The dataset from this study is held securely in coded form at ICES. While legal data sharing agreements between ICES and data providers (e.g., healthcare organizations and government) prohibit ICES from making the dataset publicly available, access may be granted to those who meet pre-specified criteria for confidential access, available at [www.ices.on.ca/DAS](http://www.ices.on.ca/DAS) (email: [das@ices.on.ca](mailto:das@ices.on.ca)).

### Editor note

The *Lancet* Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

### Funding

Canadian Institutes of Health Research, and Peter Munk Cardiac Centre.

### Declaration of interests

Dr Udell has received consulting fees or honoraria from Amgen, Boehringer Ingelheim, Novartis and Sanofi, and grant support from Boehringer Ingelheim and Janssen. All authors declare no conflicts of interest.

### Acknowledgments

This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health (MOH) and the Ministry of Long-Term Care (MLTC), and by the Ontario Health Data Platform (OHDP), a Province of Ontario initiative to support Ontario's ongoing response to COVID-10 and its related impacts. This study also received funding from a Canadian Institutes of Health Research, COVID-19 Rapid Research Funding Opportunity grant (# VR4 172736) and a Peter Munk Cardiac Centre Innovation Grant. Parts of this material are based on data and information compiled and provided by the Canadian Institute for Health Information (CIHI), Cancer Care Ontario (CCO) and Immigration, Refugees and Citizenship Canada (IRCC). The analyses, conclusions, opinions and statements

expressed herein are solely those of the authors and do not reflect those of the funding or data sources; no endorsement is intended or should be inferred. Dr. D. Lee is the Ted Rogers Chair in Heart Function Outcomes, University Health Network, University of Toronto. Dr. Austin is supported by a Mid-Career investigator award from the Heart and Stroke Foundation. Dr. McAlister is supported by an Alberta Health Services Chair in Cardiovascular Outcomes Research. Dr. Kaul is the CIHR Sex and Gender Science Chair and the Heart & Stroke Chair in Cardiovascular Research. Dr. Rochon holds the RTO/ERO Chair in Geriatric Medicine from the University of Toronto. Dr. B. Wang holds a CIFAR AI chair at the Vector Institute.

### Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.lana.2021.100146](https://doi.org/10.1016/j.lana.2021.100146).

### References

- 1 Brown KA, Stall NM, Vanniyasingam T, et al. Early impact of Ontario's COVID-19 vaccine rollout on long-term care home residents and health care workers. Vol. 2. 2021. Science briefs of the Ontario COVID-19 science advisory table.
- 2 Lee DS, Ma S, Chu A, et al. Predictors of mortality among long-term care residents with SARS-CoV-2 infection. *J Am Geriatr Soc*. 2021. <https://doi.org/10.1111/jgs.17425>.
- 3 Abrams HR, Loomer L, Gandhi A, Grabowski DC. Characteristics of U.S. nursing homes with COVID-19 cases. *J Am Geriatr Soc*. 2020;68(8):1653–1656. <https://doi.org/10.1111/jgs.16661>.
- 4 Li Y, Temkin-Greener H, Shan G, Cai X. COVID-19 infections and deaths among connecticut nursing home residents: facility correlates. *J Am Geriatr Soc*. 2020;68(9):1899–1906. <https://doi.org/10.1111/jgs.16689>.
- 5 Stall NM, Jones A, Brown KA, Rochon PA, Costa AP. For-profit long-term care homes and the risk of COVID-19 outbreaks and resident deaths. *CMAJ Can Med Assoc J*. 2020;192(33):E946–E955. <https://doi.org/10.1503/cmaj.201197>.
- 6 Shallcross L, Burke D, Abbott O, et al. Factors associated with SARS-CoV-2 infection and outbreaks in long-term care facilities in England: a national cross-sectional survey. *Lancet Healthy Longev*. 2021;2(3):e129–e142. [https://doi.org/10.1016/S2666-7568\(20\)30065-9](https://doi.org/10.1016/S2666-7568(20)30065-9).
- 7 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(3):e216315. <https://doi.org/10.1001/jamanetworkopen.2021.6315>.
- 8 Li Y, Cen X, Cai X, Temkin-Greener H. Racial and ethnic disparities in COVID-19 infections and deaths across U.S. nursing homes. *J Am Geriatr Soc*. 2020;68(11):2454–2461. <https://doi.org/10.1111/jgs.16847>.
- 9 Canadian Institute for Health Information. *Continuing Care Reporting System RAI-MDS 2.0 Output Specifications, 2010-2011*. Ottawa, ON: CIHI; 2009.
- 10 Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '16*. 2016:785–794. <https://doi.org/10.1145/2939672.2939785>. Accessed August 2016.
- 11 Friedman J, Hastie T, Tibshirani R. *The Elements of Statistical Learning*. New York: Springer series in statistics; 2001.
- 12 Lundberg S, Lee SI. A unified approach to interpreting model predictions. arXiv. 2017;preprint arXiv:1705.07874. <https://arxiv.org/abs/1705.07874>.
- 13 Patel MC, Chaisson LH, Borgetti S, et al. Asymptomatic SARS-CoV-2 infection and COVID-19 mortality during an outbreak investigation in a skilled nursing facility. *Clin Infect Dis*. 2020;71(11):2920–2926. <https://doi.org/10.1093/cid/ciaa763>.

- 14 Sun CLF, Zuccarelli E, Zerhouni EGA, et al. Predicting coronavirus disease 2019 infection risk and related risk drivers in nursing homes: a machine learning approach. *J Am Med Dir Assoc.* 2020;21(11):1533–1538. <https://doi.org/10.1016/j.jamda.2020.08.030>.
- 15 Shi SM, Bakaev I, Chen H, Trivison TG, Berry SD. Risk factors, presentation, and course of coronavirus disease 2019 in a large, academic long-term care facility. *J Am Med Dir Assoc.* 2020;21(10):1378–1383. <https://doi.org/10.1016/j.jamda.2020.08.027>.
- 16 Brown KA, Jones A, Daneman N, et al. Association between nursing home crowding and COVID-19 infection and mortality in Ontario, Canada. *JAMA Intern Med.* 2020. <https://doi.org/10.1001/jamainternmed.2020.6466>.
- 17 Dumont-Leblond N, Veillette M, Bherer L, et al. Positive no-touch surfaces and undetectable SARS-CoV-2 aerosols in long-term care facilities: an attempt to understand the contributing factors and the importance of timing in air sampling campaigns. *Am J Infect Control.* 2021. <https://doi.org/10.1016/j.ajic.2021.02.004>.
- 18 van Doremalen N, Bushmaker T, Morris DH, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N Engl J Med.* 2020;382(16):1564–1567. <https://doi.org/10.1056/NEJMc2004973>.
- 19 Elbadawy HM, Khattab A, Alalawi A, et al. The detection of SARS-CoV-2 in outpatient clinics and public facilities during the COVID-19 pandemic. *J Med Virol.* 2021. <https://doi.org/10.1002/jmv.26819>.
- 20 Belmin J, Um-Din N, Donadio C, et al. Coronavirus disease 2019 outcomes in French nursing homes that implemented staff confinement with residents. *JAMA Netw Open.* 2020;3(8):e2017533. <https://doi.org/10.1001/jamanetworkopen.2020.17533>.
- 21 Sacco G, Foucault G, Briere O, Annweiler C. COVID-19 in seniors: findings and lessons from mass screening in a nursing home. *Maturitas.* 2020;141:46–52. <https://doi.org/10.1016/j.maturitas.2020.06.023>.
- 22 Griffith GJ, Morris TT, Tudball MJ, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun.* 2020;11(1):5749. <https://doi.org/10.1038/s41467-020-19478-2>.
- 23 Sundaram ME, Calzavara A, Mishra S, et al. Individual and social determinants of SARS-CoV-2 testing and positivity in Ontario, Canada: a population-wide study. *CMAJ Can Med Assoc J.* 2021;193(20):E723–E734. <https://doi.org/10.1503/cmaj.202608>.