

Derivation and Validation of Clinical Prediction Rules for COVID-19 Mortality in Ontario, Canada

David N. Fisman,¹ Amy L. Greer,² Michael Hillmer,^{1,3,4} and R. Tuite^{1,6}

¹The Dalla Lana School of Public Health, University of Toronto, Ontario, Canada, ²The Department of Population Medicine, University of Guelph, Ontario, Canada, ³The Ontario Ministry of Health, Ontario, Canada, and ⁴The Dalla Lana School of Public Health, Ontario, Canada

Background. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is currently causing a high-mortality global pandemic. The clinical spectrum of disease caused by this virus is broad, ranging from asymptomatic infection to organ failure and death. Risk stratification of individuals with coronavirus disease 2019 (COVID-19) is desirable for management, and prioritization for trial enrollment. We developed a prediction rule for COVID-19 mortality in a population-based cohort in Ontario, Canada.

Methods. Data from Ontario's provincial iPHIS system were extracted for the period from January 23 to May 15, 2020. Logistic regression-based prediction rules and a rule derived using a Cox proportional hazards model were developed and validated using split-halves validation. Sensitivity analyses were performed, with varying approaches to missing data.

Results. Of 21 922 COVID-19 cases, 1734 with complete data were included in the derivation set; 1796 were included in the validation set. Age and comorbidities (notably diabetes, renal disease, and immune compromise) were strong predictors of mortality. Four point-based prediction rules were derived (base case, smoking excluded, long-term care excluded, and Cox model-based). All displayed excellent discrimination (area under the curve for all rules > 0.92) and calibration ($P > .50$ by Hosmer-Lemeshow test) in the derivation set. All performed well in the validation set and were robust to varying approaches to replacement of missing variables.

Conclusions. We used a public health case management data system to build and validate 4 accurate, well-calibrated, robust clinical prediction rules for COVID-19 mortality in Ontario, Canada. While these rules need external validation, they may be useful tools for management, risk stratification, and clinical trials.

Keywords. aging; clinical epidemiology; COVID-19; logistic models; SARS-CoV-2.

Since the coronavirus disease 2019 (COVID-19) pandemic was declared by the World Health Organization on March 12, 2020 [1], the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has taken a fearsome toll on global mortality. As of June 11, 2020, >400 000 deaths worldwide have been attributed to SARS-CoV-2, with many more excess deaths likely related either to infection with the virus or disruption of health systems by epidemics [2]. While most infections with SARS-CoV-2 are mild or even asymptomatic, ~20% of recognized infections are sufficiently severe to require hospitalization [3, 4]. Among those hospitalized, 10%–20% have an intensive care requirement, usually related to respiratory failure [3–5], though multiorgan system failure [6], clotting abnormalities [7], and angiogenesis [8] with resultant bleeding are increasingly recognized as severe complications of COVID-19.

Numerous studies have identified clinical factors associated with requirement for intensive care and death among those with COVID-19 infection [9–14]. Published prediction models to date have evaluated case-level factors that might predict care diagnosis, more severe disease requiring hospitalization, and poor outcomes (critical illness or death) [9]. A recent review identified 16 prediction models focused on prognosis; 14 were based on the COVID-19 epidemic in China, and the other 2 used aggregated public data from a variety of sources [9]. Existing models focused on patients in care and excluded community cases of COVID-19 who were never hospitalized. The generalizability of these rules to the North American context is unclear. Furthermore, few of these efforts included conversion of prediction models into parsimonious, simple, score-based tools that can be used easily for risk stratification in clinical settings. In the context of COVID-19, such rules might have important implications for risk stratification of patients, streamlining decisions around hospital care vs self-isolation [15], and prioritizing individuals for enrollment in clinical trials of emerging therapies (eg, convalescent plasma or antiviral drugs), as has been the case with similar tools developed for community-acquired pneumonia [16].

Ontario, Canada, had identified >30 000 virologically confirmed cases of COVID-19 in the province as of June 11, 2020

Received 3 July 2020; editorial decision 22 September 2020; accepted 30 September 2020.

Correspondence: David Fisman, MD, MPH, FRCP(C), Room 686, 155 College Street, Toronto, ON, M5T 3M7, Canada (david.fisman@utoronto.ca).

Open Forum Infectious Diseases®

© The Author(s) 2020. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
DOI: 10.1093/ofid/ofaa463

[17]. Each confirmed case is the subject of investigation by local public health authorities, who enter epidemiological, clinical, and outcome data into the Province's Integrated Public Health Information System (iPHIS). Our objective was to make use of these data to develop and validate parsimonious, sensitive, and specific prediction rules for infection-related death in individuals with COVID-19 in Ontario.

METHODS

Study Population and Data Collection

Ontario is Canada's most populous province, with a current population of 14.7 million [18]. The province identified imported COVID-19 cases from China and Iran in January and February 2020, respectively [19]; local epidemic spread of SARS-CoV-2 has been evident since late February 2020 [20]. Each of Ontario's 34 public health units is responsible for local case investigation and uploading of case information into the iPHIS data system, which is used for surveillance and case management of notifiable diseases in the province [21]. Ontario's case definition for a confirmed case requires a positive laboratory test using a validated nucleic acid amplification test, including real-time polymerase chain reaction and nucleic acid sequencing [22]. As this is a public health case management system, those entering data are not blinded to case outcome. Definitions used by public health professionals completing case records are available at https://myrnao.ca/sites/default/files/attached_files/PH%20Mgmt%20nCoV%20Shared%202020-02-06.pdf.

Information on patient characteristics, including age group (by 10-year intervals), sex, medical comorbidities, long-term care residence, health care and emergency service work, case symptoms, dates of symptom onset, testing and reporting, hospitalization and intensive care admission, and mortality, was available for cases. Approximately 80% of all deaths during the Ontario COVID-19 epidemic have occurred in long-term care facilities [23], and there has been limited transfer of long-term care residents to intensive care units [20].

Statistical Analysis

The recently published TRIPOD guidelines have established best practices for the construction, validation, and presentation of clinical prediction models. This work was performed in adherence to the 21-point TRIPOD checklist [24]. We randomly assigned cases into derivation and validation sets. Univariable logistic regression was used to identify factors associated with mortality in the derivation group. Continuous variables were dichotomized to facilitate score generation and ease of application in clinical settings. When a factor was found to be protective, the covariate evaluated was *absence* of the factor, so that resultant odds ratios were >1 .

Risk factors significant at $P < .2$, or which were thought a priori to confer important increases in risk (age and sex), were included in model building using a forward stepwise selection

algorithm, with covariates selected for $P < .05$, and retained in the model for $P < .15$. We did not include interaction terms in an effort to keep the final prediction rule as simple as possible. The final regression model was transformed to a point-based rule, with each regression coefficient divided by half of the smallest coefficient and rounded to the nearest integer to obtain weighted values. Risk scores were calculated by summing the individual point values of all applicable risk factors. Risk of death can then be approximated from a graph of model-predicted probability vs calculated score (Figure 1) using the relation $P = 1/(e^{-(I + CS)} + 1)$, where S is the individual's score, C is the prediction rule's coefficient in a logit model using score as a predictor of death, and I is the intercept from the same model.

The discriminatory ability of the prediction rule in the derivation group was quantified through the area under the receiver operating characteristic curve (ROC AUC), with 95% confidence intervals estimated through 1000 bootstrap replicates. Calibration was assessed visually and using the

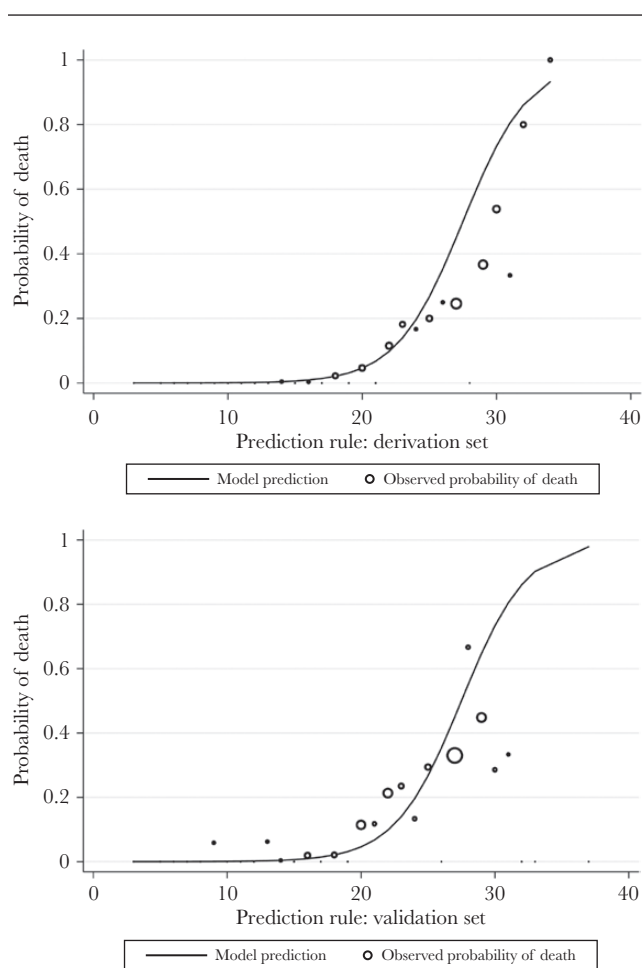


Figure 1. Observed and predicted risk of death by score, base-case rule. Plot of predicted probability of death (y-axis) by model score (x-axis) for base case prediction rule. Curve represents model predictions; circles represent observed proportion who died. Circle size is proportionate to number of deaths at a given score. Top panel: derivation set; bottom panel: validation set.

Hosmer-Lemeshow test for goodness of fit, which evaluates expected and observed probabilities in population deciles [25]. As both calibration and discrimination were excellent in the validation set, recalibration was not performed.

Survival Analytic Approach and Alternate Rules

Some analysts have expressed concern that failure to account for right censoring could lead to bias in COVID-19 clinical prediction rules [9]. As such, we created a second prediction rule using Cox proportional hazards analysis by identifying factors associated with increased hazard of death using the same selection algorithm as applied to the logistic model described above. Log-transformed hazard ratios were converted to point scores using the approach described above. Discriminative ability of the rule was evaluated using Harrell's C-statistic after constructing a Cox proportional hazards model with the score as the sole covariate in both the derivation and validation sets. ROC analysis and score calibration were performed using the Cox model-derived score as a predictor in a logistic model.

Smoking status emerged as a protective effect in our base case prediction model; this is likely to be controversial with some users. Furthermore, it might be argued that the known high mortality associated with COVID-19 in long-term care settings favors creation of a rule for non-long-term care residents. As such, we made additional rules that excluded smoking status and excluded long-term care residents, using the approach described above.

Sensitivity Analyses

In the base case, models were built using only observations from individuals with complete data; we tested the robustness of our models by evaluating the discriminative ability and calibration of rules in data sets in which missing fields were replaced at random and in data sets where an attribute was assumed not present if a field was left blank (eg, if an individual had no record of presence or absence of cardiac disease, they were assumed not to have cardiac disease). All analyses were performed using Stata, version 14.0 (Stata Corporation, College Station, TX, USA). The study was approved by the research ethics board of the University of Toronto.

Patient Consent Statement

Due to the retrospective nature of the study, which involved precollected data assembled as a component of epidemic response, and the absence of identifying personal health information in the data set, written informed consent by subjects was not required by the research ethics board and was not obtained.

Conflicts of Interest

None of the authors has any conflict of interest associated with the study or with the publication of this work.

RESULTS

Of 21 922 COVID-19 cases reported between January 23 and May 15, 2020, 57% were female, and the median age (interquartile range [IQR]) was 55 (35–75) years. The median time from symptom onset to case reporting (IQR) was 5 (4–10) days. Fourteen percent of cases were residents of long-term care facilities; 17% were health care workers. Thirteen percent of cases were hospitalized; 2% had record of intubation and/or mechanical ventilation, and case fatality was 8%, which may have reflected low hospitalization rates among residents of long-term care facilities.

Individuals assigned to the derivation set were significantly more likely to be smokers, less likely to have a history of chronic liver disease, and less likely to die (Table 1). A total of 1734 individuals with complete data on all candidate predictors were used to construct the base case prediction model.

Derivation of the Prediction Rule

In univariable analyses, death was associated with a broad array of demographic characteristics and comorbid conditions. No association was seen between risk of death and mean neighborhood income or asthma, which were not included in subsequent model building (Table 2). As age was provided as ordinal 10-year age groupings (0–9, 10–19, 20–29, etc.), the age coefficient in models represents increased risk per increase in (age/10).

Using a forward selection algorithm, we identified 7 independent predictors of death in the derivation group: age, long-term care residence, a history of renal disease, diabetes, chronic obstructive pulmonary disease, immune compromise, and non-smoking. (Table 3).

The point-based prediction rule was well calibrated between quantiles of observed and expected risk (Hosmer-Lemeshow $\chi^2 = 1.58$; $P = .90$) in the derivation group and discriminated extremely well between those who did and did not die (ROC AUC in the derivation group, 0.95; 95% CI, 0.91–0.96). The median score (IQR) was 13 (6) for survivors and 25 (6) for those who died ($P < .001$ by the Wilcoxon rank-sum test). The rule displayed good calibration to outcomes in the 1738 individuals with complete data in the validation set (Hosmer-Lemeshow $\chi^2 = 9.16$; $P = .16$), as well as excellent discrimination (AUC, 0.92; 95% CI, 0.89–0.94) (Figures 1 and 2).

Alternate Prediction Rules

Three alternate rules (a model based on a Cox proportional hazards model, a logistic model excluding smoking status, and a model with long-term care residents excluded) were created. These models had excellent discrimination. We found statistical evidence of poor calibration of the model that excluded long-term care residents in the validation set ($P = .04$). The Harrell's C-statistic for a Cox model including age, male sex, diabetes, chronic obstructive pulmonary disease, and immune

Table 1. Characteristics of Confirmed COVID-19 Cases in Ontario, Canada, Through May 15, 2020

Covariate (% Reported ^a)	Overall (%) or Median (IQR)	Derivation (%) or Median (IQR)	Validation (%) or Median (IQR)	P Value
All	21 922 (100)	10 957 (50)	10 965 (50)	
Age (100)	55 (40)	55 (40)	55 (40)	.63
FSA income, ^b \$ (96)	64 869 (26 402)	64 986 (26 937)	64 869 (26 402)	.73
Male gender (100)	9389 (43)	4708 (43)	4681 (43)	.68
Time from symptom onset to report (58)	5 (6)	5 (6)	5 (6)	.81
Long-term care resident (46)	3102 (14)	1539 (14)	1563 (14)	.50
Outbreak-associated case ^c (100)	9438 (43)	4772 (44)	4666 (43)	.14
Health care worker (47)	3780 (17)	1888 (17)	1892 (17)	.55
Homeless shelter worker (27)	106 (0.4)	61 (0.6)	45 (0.4)	.13
Homeless (29)	226 (1)	102 (0.9)	124 (1)	.11
Smoker (30)	515 (2)	285 (3)	230 (2)	.01
Pregnant or postpartum (23)	91 (0.4)	58 (0.5)	45 (0.4)	.21
Comorbidity history				
Anemia or hemoglobinopathy (23)	370 (2)	177 (2)	193 (2)	.42
Chronic liver disease (23)	94 (0.4)	38 (0.3)	56 (0.5)	.06
Renal disease (21)	358 (2)	185 (2)	173 (2)	.55
Diabetes (26)	1294 (6)	623 (6)	671 (6)	.22
Chronic obstructive pulmonary disease (23)	267 (1)	134 (1)	133 (1)	.97
Asthma (25)	880 (4)	409 (4)	390 (4)	.58
Cardiovascular disease (25)	2032 (9)	979 (9)	1053 (10)	.15
Malignancy (24)	460 (2)	211 (2)	249 (2)	.06
Immune compromise (25)	318 (1)	162 (1)	156 (1)	.69
Tuberculosis (22)	52 (0.2)	32 (0.2)	20 (0.2)	.10
Obesity (20)	295 (1)	137 (1)	158 (1)	.16
Outcomes				
Hospitalized (100)	2779 (13)	1355 (12)	1424 (13)	.17
Record of intubation and/or mechanical ventilation (20)	408 (2)	195 (2)	213 (2)	.31
Died (46) ^d	1825 (8)	862 (8)	963 (9)	.02

Proportions were compared with the chi-square test, and continuous variables were compared with the Wilcoxon rank-sum test.

Abbreviations: COVID-19, coronavirus disease 2019; FSA, forward sortation area; IQR, interquartile range.

^aDefined as total number of individuals with nonmissing information in that field, divided by total cohort size (21 922). Total number with complete information is (percent reported × 21 922)/100.

^bBased on mean after tax income in Canadian dollars by forward sortation area (2016 Canadian Census). Forward sortation area is defined by the first 3 alphanumeric characters of a Canadian postal code.

^cDefined as case or cases with outbreak number signifying part of an outbreak investigation by a public health unit.

^dA total of 10 053 individuals (46% of the cohort) had information entered in the "outcome" field, which included death, recovery, "ill," "unknown," residual effects, and pending outcomes.

compromise was 0.93 in the derivation set and 0.91 in the validation set. The number of individuals with complete data available for model construction, other fit statistics, and c-statistics for AUC, as well as values of the model intercept and smallest logit model coefficient (for calculation of death probability), are presented in [Table 3](#) and presented graphically in the [Supplementary Data](#).

Sensitivity Analyses

We re-evaluated all 4 prediction rules in data sets in which missing variables were assumed to not be present and in which missing variables were replaced randomly. Discriminative ability remained good for both randomly replaced data sets (ROC curve AUC, 0.84–0.90 for missing observations replaced with zeroes; AUC, 0.79–0.83 for missing observations replaced randomly). The large number of observations in data sets with all missing observations replaced (n = 21 922) resulted in statistically significant differences between observed and expected

mortality probabilities ($P < .001$ for all analyses by Hosmer-Lemeshow test), but visual inspection suggested that calibration of rules remained very good ([Supplementary Data](#)).

DISCUSSION

Accurate prediction of mortality from COVID-19 has a number of potential applications, including rational decision-making for hospital admission, prioritization of high-risk individuals for inclusion in trials of novel therapeutic agents, and identification of high-risk individuals for policy purposes (eg, to inform decisions around risks and benefits of remote work). We demonstrate here that COVID-19 mortality in identified cases can be predicted with remarkable accuracy based on the limited, readily available demographic and chronic health information available in public health line lists. The large number of COVID-19 cases that have occurred in Ontario provided sufficient statistical power for both model derivation and validation

Table 2. Univariable and Multivariable Analyses and Point Score Derivation, Base Case Prediction Rule

Covariate	Univariable OR (95% CI)	PValue	Multivariable OR (95% CI)	Logit	Points
Age (per 10-y increment)	3.48 (3.28–3.70)	<.001	2.42 (1.78–3.29)	0.88	2
Low income ^a	1.04 (0.92–1.18)	.55	—	—	—
Male gender	1.13 (1.02–1.25)	.02	—	—	—
Time from symptoms to diagnosis ≤3 d ^b	1.27 (1.14–1.42)	<.001	—	—	—
Long-term care resident	22.62 (19.08–26.83)	<.001	6.24 (2.95–13.21)	1.83	4
Outbreak-associated case	9.15 (8.10–10.33)	<.001	—	—	—
Non-health care worker ^c	30.56 (15.77–59.22)	<.001	—	—	—
Non-homeless shelter worker ^c	5.79 (0.80–41.96)	.08	—	—	—
Nonhomeless ^c	2.31 (0.94–5.71)	.07	—	—	—
Nonsmoker ^c	1.65 (0.98–2.77)	.06	6.86 (0.73–64.27)	1.93	4
Pregnant or postpartum	No deaths	—	—	—	—
Comorbidity history					
Anemia or hemoglobinopathy	5.08 (3.68–7.02)	<.001	—	—	—
Chronic liver disease	6.06 (3.50–10.46)	<.001	—	—	—
Renal disease	9.85 (7.31–13.26)	<.001	2.37 (0.97–5.77)	0.86	2
Diabetes	6.49 (5.22–8.06)	<.001	2.19 (1.08–4.42)	0.78	2
Chronic obstructive pulmonary disease	11.22 (8.14–15.44)	<.001	3.26 (1.15–9.26)	1.18	3
Asthma	1.01 (0.71–1.44)	.96	—	—	—
Cardiovascular disease	11.38 (9.12–14.20)	<.001	—	—	—
Malignancy	6.36 (4.80–8.44)	<.001	—	—	—
Immune compromised	4.12 (2.94–5.79)	<.001	3.56 (1.12–11.35)	1.27	3
Tuberculosis	0.88 (0.21–3.70)	<.001	—	—	—
Obesity	2.63 (1.78–3.89)	<.001	—	—	—

Abbreviations: FSA, forward sortation area; OR, odds ratio.

^aResidence in FSA in lowest quartile of income.

^bLowest quartile lag between symptoms and diagnosis.

^cNonexposure status evaluated as risk factor to maintain positive covariate.

without requiring bootstrap resampling. The discriminative ability of our rules (as reflected in AUC > 0.9 in both the derivation and validation sets) places them among the upper tier of

current COVID-19 prediction rules; the parsimoniousness of these rules and their conversion to an easy-to-calculate point score allows easy incorporation into clinical care. Our rules are

Table 3. Base Case and Alternate Clinical Prediction Rules

Covariate	Rule 1: Base Case	Rule 2: Cox Model-Based ^a	Rule 3: Nonsmokers Excluded	Rule 4: Long-term Care Residents Excluded
Age/10	2	3	3	2
Male sex	—	2	—	—
Renal disease	2	—	2	3
Immune compromised	3	4	5	4
Diabetic	2	4	3	2
COPD	3	3	3	—
Cardiovascular disease	—	—	2	4
Long-term care resident	5	—	7	—
Nonsmoker	5	—	—	—
Time from symptoms to diagnosis ≤3 d	—	—	—	2
Maximum points	40	40	50	40
No. in derivation set (validation set)	1734 (1738)	2348 (2340)	1893 (1874)	2285 (2271)
Smallest logit model coefficient ^a	0.36	0.34	0.25	0.52
Model intercept ^a	−9.81	−9.99	−8.33	−12.51
AUC in derivation set (validation set)	0.95 (0.92)	0.93 (0.91)	0.95 (0.92)	0.92 (0.91)
Hosmer-Lemeshow goodness-of-fit test P value in derivation set (validation set)	0.85 (0.20)	0.50 (0.24)	0.99 (0.40)	0.59 (0.04)

Hosmer-Lemeshow test based on deciles of risk score.

Abbreviations: AUC, area under the receiver operating characteristics curve; COPD, chronic obstructive pulmonary disease.

^aCan be used to calculate probability of death as per text.

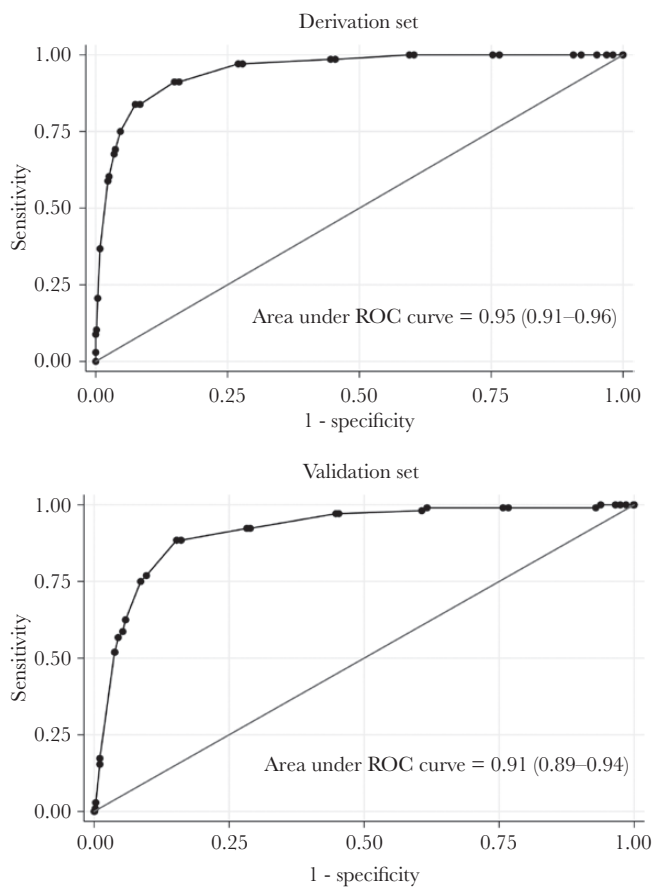


Figure 2. Receiver operator characteristic curve, base-case rule. Sensitivity of rule (y-axis) is plotted against false-positive rate (1-specificity, x-axis) for different positivity criteria available from score. Confidence intervals for area under the curve were derived via bootstrapping. Top panel: derivation set; bottom panel: validation set.

also, we believe, unique in that they are based on an entire population of identified COVID-19 cases, and not restricted to individuals who have accessed the health care system.

Many of our predictors (age and comorbidities) could have been anticipated based on the established epidemiology of COVID-19 [26–28]. The identification of nonsmoking as a predictor of mortality is likely to be controversial, and it is for this reason that we derived alternate rules that exclude nonsmoking. Apparent protective effects of smoking against COVID-19 acquisition [29] as well as under-representation of smokers among COVID-19 patients have been noted by others [30]. However, other investigators have suggested higher risk of progression of COVID-19 in smokers [30, 31] and increased density of ACE-2 (a viral receptor) in the lungs of smokers [32], such that apparent protective effects might result from selection bias; individuals predisposed to very mild COVID-19 infection as a result of young age or good general health might be over-represented among those tested for COVID-19 due to smoking-related health concerns like cough. Regardless, a noncausal association with risk may still

be useful for clinical prediction. If this association reflects peculiarities of Ontario’s approach to COVID-19 testing, we expect that it may not be generalizable to other jurisdictions that test more widely.

Similarly, the strong effect of long-term care residence on mortality is unsurprising, given the high fraction of long-term care deaths seen during the Canadian COVID-19 epidemic to date [23]. As such we created alternate rules that exclude smoking and long-term care residence; these rules can be used in place of our base case rule, as they have similar discriminative ability. Lastly, to avoid biases that might be introduced by right-censoring (ie, lack of mortality in individuals in the study cohort as a result of insufficient follow-up time) we derived an additional rule using survival methods, which also performed well. There was substantial overlap between all 4 prediction rules in included covariates: Notably, age, diabetes, and immune compromise were included in all 4 rules we derived, and renal or chronic obstructive pulmonary disease was included in 3 of 4 rules.

Our analysis had many limitations. The use of a public health record system not explicitly designed as a research tool means that we lack laboratory and radiological results that have been useful in other prediction models [10, 27]. Furthermore, missing data were a significant limitation of our data set, although our models appeared robust even with random replacement of predictors and outcomes that should bias associations toward the null. In that sense, the ability to derive simple, accurate, and parsimonious rules, which perform well in split-halves validation, despite limitations in our data set, may suggest generalizability of application outside Ontario. We hope that other groups will evaluate our rules in other settings.

In summary, we developed and internally validated a prediction rule for COVID-19 mortality using a large and detailed public health line list in the Canadian province of Ontario. The rule was well calibrated, discriminated well, and was robust in sensitivity analyses to assess the impact of missing information on predictor variables. If externally validated, this rule might facilitate decision-making during future epidemic waves.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases online*. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Financial support. The research was supported by a grant to D.N.F. from the Canadian Institutes for Health Research (2019 COVID-19 rapid researching funding OV4-170360).

Potential conflicts of interest. None of the authors has any conflict of interest to declare in connection to this work.

References

1. World Health Organization. Statement on the meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). 2020. Available at: <https://www.who.int/news/item/23-01-2020-statement-on-the-meeting-of-the-international-health->

- regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov). Accessed 14 October 2020.
2. Johns Hopkins University. Coronavirus resource tracker. 2020. Available at: <https://coronavirus.jhu.edu/data/new-cases>. Accessed 20 June 2020.
 3. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*. 2020; 323:1061–9.
 4. Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020; 382:1708–20.
 5. Simpson CE, Chen JY, Damico RL, et al. Cellular sources of IL-6 and associations with clinical phenotypes and outcomes in PAH. *Eur Respir J*. 2020; 1901761.
 6. Zhang JJ, Dong X, Cao YY, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy*. 2020; 75:1730–41.
 7. Nahum J, Morichau-Beauchant T, Daviaud F, et al. Venous thrombosis among critically ill patients with coronavirus disease 2019 (COVID-19). *JAMA Netw Open* 2020; 3:e2010478.
 8. Ackermann M, Mentzer SJ, Jonigk D. Pulmonary vascular pathology in Covid-19. Reply. *N Engl J Med* 2020; 383:888–9.
 9. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of COVID-19 infection: systematic review and critical appraisal. *BMJ* 2020; 369:m1328.
 10. Bai X, Fang C, Zhou Y, et al. Predicting COVID-19 malignant progression with AI techniques. medRxiv 2020.03.20.20037325 [Preprint]. 12 September 2020. Available at: <https://doi.org/10.1101/2020.03.20.20037325>. Accessed 14 October 2020.
 11. Yu T, Cai S, Zheng Z, et al. Association between clinical manifestations and prognosis in patients with COVID-19. *Clin Ther*. 2020; 42:964–72.
 12. Liang W, Liang H, Ou L, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med*. 2020; 180:1081–9.
 13. Wu G, Yang P, Xie Y, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study. *Eur Respir J*. 2020; 56:2001104.
 14. Zhang S, Guo M, Duan L, et al. Development and validation of a risk factor-based system to predict short-term survival in adult hospitalized patients with COVID-19: a multicenter, retrospective, cohort study. *Crit Care* 2020; 24:438.
 15. Wollenstein-Betech S, Cassandras CG, Paschalidis IC. Personalized predictive models for symptomatic COVID-19 patients using basic preconditions: hospitalizations, mortality, and the need for an ICU or ventilator. *Int J Med Inform* 2020; 142:104258.
 16. Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997; 336:243–50.
 17. Government of Ontario. How Ontario Is Responding to COVID-19. Toronto, Canada: Queen's Printer for Ontario; 2020. Available at: <https://www.ontario.ca/page/how-ontario-is-responding-covid-19#section-0>. Accessed 10 June 2020.
 18. Statistics Canada. Population estimates, quarterly. Table: 17-10-0009-01 (formerly CANSIM 051-0005). 2020. Available at: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901>. Accessed 29 May 2020.
 19. Neilsen K. A Timeline of the Novel Coronavirus in Ontario. Toronto, Canada: Global News; 2020. Available at: <https://globalnews.ca/news/6859636/ontario-coronavirus-timeline/>. Accessed 29 May 2020.
 20. Tuite AR, Greer AL, De Keninck S, Fisman DN. Risk for COVID-19 resurgence related to duration and effectiveness of physical distancing in Ontario, Canada. *Ann Intern Med*. 2020; M20–2945.
 21. Deeks SL, Lim GH, Walton R, et al. Prolonged pertussis outbreak in Ontario originating in an under-immunized religious community. *Can Commun Dis Rep* 2014; 40:42–9.
 22. Public Health Ontario. Case Definition—Novel Coronavirus (COVID-19). Toronto, Canada: Public Health Ontario; 2020. Available at: http://www.health.gov.on.ca/en/pro/programs/publichealth/coronavirus/docs/2019_case_definition.pdf. Accessed 14 October 2020.
 23. Fisman DN, Bogoch I, Lapointe-Shaw L, McCready J, Tuite AR. Risk factors associated with mortality among residents with coronavirus disease 2019 (COVID-19) in long-term care facilities in Ontario, Canada. *JAMA Netw Open* 2020; 3:e2015957.
 24. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; 162:55–63.
 25. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York, New York: Oxford University Press; 1989.
 26. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. Vital surveillances: the epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China Centers Dis Control Weekly* 2020; 2:113–22.
 27. Hou W, Zhang W, Jin R, et al. Risk factors for disease progression in hospitalized patients with COVID-19: a retrospective cohort study. *Infect Dis (Lond)* 2020; 52:498–505.
 28. Caramelo F, Ferreira N, O'liveiros B. Estimation of risk factors for COVID-19 mortality - preliminary results. medRxiv 2020.02.24.20027268 [Preprint]. 25 February 2020. Available at: <https://doi.org/10.1101/2020.02.24.20027268>. Accessed 14 October 2020.
 29. de Lusignan S, Dorward J, Correa A, et al. Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: a cross-sectional study. *Lancet Infect Dis* 2020; 20:1034–42.
 30. Gonzalez-Rubio J, Navarro-Lopez C, Lopez-Najera E, et al. Cytokine release syndrome (CRS) and nicotine in COVID-19 patients: trying to calm the storm. *Front Immunol* 2020; 11:1359.
 31. Hamer M, Kivimaki M, Gale CR, Batty GD. Lifestyle risk factors, inflammatory mechanisms, and COVID-19 hospitalization: a community-based cohort study of 387 109 adults in UK. *Brain Behav Immun*. 2020; 87:184–7..
 32. Leung JM, Sin DD. Smoking, ACE-2, and COVID-19: ongoing controversies. *Eur Respir J*. 2020; 56:2001759.