Studying the social determinants of COVID-19 in a data vacuum

Kate H. Choi
Patrick Denice
Michael Haan
Anna Zajacova
Department of Sociology, University of Western Ontario

PWP-CCPR-2020-001

May 11, 2020

California Center for Population Research
On-Line Working Paper Series

Studying the social determinants of COVID-19 in a data vacuum

Kate H. Choi[1]

Patrick Denice[1]

Michael Haan[1]

Anna Zajacova[1]

[1]Department of Sociology, University of Western Ontario

May 10, 2020

Running title: COVID-19 data vacuum

Keywords: COVID-19, neighborhood contexts, health inequality, racial inequality

Abstract

The Canadian government has no plans to release data on the race or socioeconomic status of COVID-19 patients. Therefore, whether COVID-19 is disproportionately affecting certain socio-demographic groups in Canada is unknown. We fill this data void by merging publicly available COVID-19 data with tabular census data to identify risk factors rendering certain geographic areas more vulnerable to COVID-19 infections and deaths. We combine insights obtained from this analysis with information on the socio-demographic profiles of smaller geographic units to predict and display the incidence of COVID-19 infections and deaths in these locales. Like in the U.S., COVID-19 has disproportionately affected black and immigrant communities in Canada. COVID-19 death tolls are also higher in Canadian communities with higher shares of older adults.

Introduction

Black, low-income, and immigrant communities in the U.S. are particularly vulnerable to COVID-19 (Chin et al. 2020; Thebault et al. 2020). Whether this is also true in Canada is largely unknown because researchers do not have access to data about the race/ethnicity or socioeconomic status of COVID-19 patients (Ramos 2020; Siddiqi et al., 2020). Despite repeated pleas for such data from social scientists, Canadian government officials have responded that they have no plans to make these data available because *all groups are equally important* (Osman, 2020; Mulligan et al., 2020).

Canadian researchers are currently relying on two data sources. One comes from the Public Health Agency of Canada (PHAC). These data report the daily infection and death counts for 86 health regions, but they are not tallied separately by race, socioeconomic status, or demographic profile. Furthermore, each health region administers and disperses health care resources to approximately 420,000 residents; thus, they are too large geographically and too socially heterogeneous to capture the impact of COVID-19 on local communities. The second is crowdsourced data, which relies on self-reports of COVID-19 cases by postal code (Jones, 2020). Postal code areas are small enough to represent communities. These data, however, can miss a significant portion of COVID-19 cases by relying on self-reports to internet sites. These data identify vulnerable communities, but they only consider population density and percent of older residents (70+), ignoring the community's racial and socioeconomic composition.

Whether or not black, immigrant, and low-income communities in Canada will be disproportionately affected by COVID-19 is *a priori* unclear. Like in the U.S., racial minorities and immigrants in Canada are more economically disadvantaged than whites

with comparable levels of education (Attello et al., 2010). Health inequality also exists along socioeconomic lines in Canada (Hajizadeh et al., 2016). There are, however, reasons to believe that the pattern of the COVID-19 spread will differ across the two countries. Canada has universal health insurance coverage; the U.S. does not. Black Canadians report health comparable to whites unlike African Americans who have a health disadvantage (Veenstra and Patterson, 2016).

In the absence of data about the socio-demographic profiles of COVID-19 patients, we link aggregate level COVID-19 data with tabular census data detailing the socio-demographic profile of communities. We identify the risk factors for COVID-19 infections and death rates in health regions. Coefficients from this analysis are combined with information about the socio-demographic traits of local communities to predict their counts of COVID-19 infections/deaths. These analyses reveal that black and immigrant communities in Canada are disproportionately impacted by COVID-19.

Methods

Data and Sample

We obtained the daily counts of COVID-19 infections and deaths through May 5 for each health region, which was released by the PHAC (Berry et al. 2020). We restrict our sample to 82 (out of 86) health regions. We exclude 3 regions in Yukon, the Northwest Territories, and Nunavut because they are sparsely populated and report very few COVID-19 cases. We exclude the Southwestern Health Region in Ontario because it was not identified consistently across our data sources.

Tabular data from the 2016 Canadian Census reports the socio-demographic characteristics of census subdivisions (CSDs) and divisions (CDs). These geographic areas

are smaller and represent local communities better than health regions. CDs are equivalent to U.S. counties; CSDs are nested inside CDs and usually denote a municipality, representing cities or towns. Statistics Canada suppresses information about CSDs with fewer than 5,000 residents. Therefore, in our study, a community is defined as either a CSD with more than 5,000 residents or remaining areas within the CD outside of populous CSDs. There are 967 communities in our study: 707 CSDs and 260 areas outside populous CSDs. The average population size in a community is roughly 6,900.

We use these socio-demographic data to describe the populations in the larger geographic areas (health regions) and smaller communities. Because CSDs/CDs are nested inside heath regions, we aggregate socio-demographic characteristics from the CSDs/CDs to describe the population composition of health regions.

Measures

Our outcomes are cumulative counts of *infections* and *deaths* through May 5, 2020 in each health region. Our covariates are *percent black*, *median after-tax income*, *percent foreign-born*, *percent 65 years and older*, and *population density*, which are all known risk factors for COVID-19 (Thebault, 2020; Yancy, 2020). These measures are computed for health regions and communities. Other measures capturing socio-demographic composition (e.g., percent East Asian, percent college-educated) are excluded due to multicollinearity or data quality issues.

Analytical strategy

The first part identifies the risk factors for COVID-19 in health regions. We estimate negative binomial regression models and determine the association between the racial, socioeconomic, and demographic traits of health regions and COVID-19

infection/death counts. We select this model over Poisson regression models because COVID-19 counts are over-dispersed. Community characteristics are presented as percentages, but we also estimated models with standardized covariates and obtained similar results.

The second part predicts the relative infection and death counts in each community using the coefficients obtained from the negative binomial regression models in part 1 and information about the socio-demographic composition of these communities. The predicted counts were used as weights to partition the actual number of COVID-19 infections/deaths in a health region across its constituent communities.

Results

Descriptive results

Figure 1 presents the spatial distribution of black, immigrant, and low-income communities in Canada. Black and immigrant communities are concentrated in the three largest metropolitan areas: Montreal, Toronto, and Vancouver. Noteworthy clusters of black communities outside of these cities are in Northeast Alberta and Nova Scotia. Median after-tax income is also higher in these metropolitan areas.

Figure 1 goes here.

Figure 2 shows the spatial distribution of COVID-19 infections and deaths across the 82 health regions in our analysis. COVID-19 rates are high in the three largest cities in Canada. Outside of them, infection rates are high on nearly all health regions in the center of Quebec as well as the Vancouver Coast and Calgary health regions. The spatial distribution of COVID-19 deaths is similar to that of infections.

Figure 2 goes here.

Multivariate analysis

Panel A in Table 1 presents the results from negative binomial regression models predicting the COVID-19 infection incidence rates for the 82 health regions. COVID-19 infection rates are significantly higher in health regions with higher shares of black residents. A 1 percentage point increase in the share of black residents in a health region is associated with the doubling of COVID-19 infection rates, net of other characteristics. COVID-19 infection rates are higher in health regions with higher percentages of foreign-born residents. A 1 percentage point increase in the share of foreign-born residents is associated with a 3% rise in COVID-19 infection rates, holding other characteristics constant. Differences in COVID-19 infection rates by income level, population density, and age structure are minimal.

Table 1 goes here.

Panel B presents results from analyses predicting COVID-19 deaths for the 82 health regions. The results for COVID-19 deaths echo those for infections. For example, with a 1 percentage point increase in share of black residents, COVID-19 death rates increase by a factor of 2.1 times, net of other characteristics. Unlike for infections, percentage of residents 65 years and older is positively associated with the death rate. A 1 percentage point increase in the share of older adults is associated with a 26% increase in COVID-19 death rates, holding other variables constant.

Predicting COVID-19 infections and deaths in communities

Figure 3 displays the predicted spread of COVID-19 infections and deaths across the 967 pseudo-CSDs. A comparison of Figures 2 and 3 reveals that considerable heterogeneity exists in the spread of the COVID-19 within health regions, particularly those in the West. The City of Vancouver is potentially more vulnerable than surrounding pseudo-CSDs. The same is also true

of Northeast Alberta with large numbers of migrant workers who reside in camps while working in its oil sands. In Saskatchewan, these maps reveal pronounced urban/rural differences. Regina, the provincial capital, has been afflicted more by COVID-19 than elsewhere in the province. Finally, in Manitoba, communities near the border with Minnesota have been affected more by the pandemic than those located elsewhere.

Figure 3 goes here.

Conclusion

The absence of data containing information about the socio-demographic profiles of COVID-19 patients prevents researchers from determining whether the pandemic is disproportionately affecting certain socio-demographic groups in Canada. We filled this void by pooling COVID-19 data with tabular census data to (a) identify the risk factors associated with COVID-19's spread and (b) predict and illustrate the incidence of COVID-19 infections and deaths across smaller geographic communities.

Black communities in Canada have been disproportionately impacted by COVID-19. This may explain why places like Montreal, with large numbers of black immigrants, have emerged as Canada's COVID-19 epicenters. These results are surprising given that discrimination against blacks is thought to be less severe in Canada and the outcomes of many black Canadians, including health, are similar to those of their white counterparts (Veenstra and Patterson, 2016). Future work should pay greater attention to the experiences of black Canadians.

Immigrant communities are also particularly vulnerable to COVID-19. This pattern may partially emerge because over 90% of immigrants in Canada settle in cities with high population densities (Statistics Canada, 2013). But population density does not fully explain this relationship. The high vulnerability of immigrant communities may also emerge due to labor

market discrimination and resulting exclusion of immigrants from occupations that allow them to work from home.

The relationship between median income and COVID-19 infection/death rates is not statistically significant, independent of the health region's racial and demographic composition. This result is consistent with findings by Borjas (2020) for New York City. It may reflect the fact that individual, not community, income shapes exposure to COVID-19. Future work should test this hypothesis.

COVID-19 death rates are higher in communities with higher shares of adults ages 65 years and older. Differences in infection rates by the share of older adults are negligible. The fact that age affects deaths, but not infections, is unsurprising given findings that COVID-19 infections are more lethal among the elderly (Liu et al., 2020).

Finally, we show that collecting COVID-19 data at larger units of geography may mask important heterogeneity within these areas. This is especially true in western Canada, where health authorities have been recently consolidated into larger health regions serving more variegated populations. This consolidation potentially masks urban/non-urban differences in British Columbia and Saskatchewan, higher infection/death rates in towns bordering the U.S., and the impact of COVID-19 in industries that employ large numbers of immigrant workers and house them in group quarters.

Our study is not without limitations. First, we rely on aggregate data about the socio-demographic traits of communities and infer individual traits based on the characteristics of their community of residence. Second, our predictions assume that the risk factors associated with COVID-19 infections/deaths do not vary across communities within health regions. Third,

reliance on publicly available tabular data precludes us from considering several potential risk factors of COVID-19, including percent of essential workers.

Nonetheless, we make important contributions to COVID-19 population-health research by highlighting the importance of collecting data about the socio-demographic profiles of COVID-19 patients for smaller geographic regions. Our findings also underscore the heightened vulnerability of black and immigrant communities in Canada and suggest that containing the spread of COVID-19 may require targeting resources (e.g., testing) to these communities.

REFERENCES

Attewell, P., Kasinitz, P., & Dunn, K. (2010). Black Canadians and black Americans: racial income inequality in comparative perspective. *Ethnic and Racial Studies*, *33*, 3, 473-495. DOI: 10.1080/01419870903085883

Borjas, G. (2020). Demographic Determinants of Testing Incidence and COVID-19 Infections in New York City Neighborhoods. *NBER Working Paper* No. 26952.

Berry, I., J. Soucy, Tuite, A. & Fisman, D. (2020). Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. *Canadian Medican Association Journal, 192,* 15, E420. https://doi.org/10.1503/cmaj.75262

Chin, T., R. Kahn, R. Li, J. Chen, N. Krieger, C. Buckee, S. Balsari, & Kiang, M. (2020). U.S. county-level characteristics to inform equitable COVID-19 response. *MedRxiv* https://doi.org/10.1101/2020.04.08.20058248.

Dean, J. & K. Wilson. (2009). 'Education? It is irrelevant to my job now. It makes me very depressed …': exploring the health impacts of under/unemployment among highly skilled recent immigrants in Canada. *Ethnicity & Health*, *14*, 2, 185-204. DOI:10.1080/13557850802227049

Hajizadeh M, Mitnitski A, & Rockwood K. (2016). Socioeconomic gradient in health in Canada: Is the gap widening or narrowing? *Health Policy, 120*, 9, 1040-50. doi: 10.1016/j.healthpol.2016.07.019

Jones, A. (2020, March 27). Canadian university students create map of self-reported potential COVID-19 cases. *CTV News.* https://www.ctvnews.ca/sci-tech/canadian-university-students-create-map-of-self-reported-potential-covid-19-cases-1.4872245

Liu, K., Y. Chen, R. Lin, & Han, K. (2020). “Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients”. *The Journal of Infection.* 10.1016/j.jinf.2020.03.005

Mulligan, K., J. Rayner, and Oyne Nnorom. (2020, April 30). Race based data urgently needed during the pandemic. *The Conversation.* https://theconversation.com/race-based-health-data-urgently-needed-during-the-coronavirus-pandemic-136822.

Osman, L. (2020, May 6). Canada still only considering gathering race-based COVID-19 data. *CTV News.* https://www.ctvnews.ca/health/coronavirus/canada-still-only-considering-gathering-race-based-covid-19-data-1.4927648

Ramos, H. (2020, April 14). Being Counted in Canada’s Coronavirus Data: Stronger data collection is essential to understand how COVID-19 impacts immigrants across Canada. *New Canadian Media: The Pulse of Immigrant Canada.* https://newcanadianmedia.ca/being-counted-in-canadas-coronavirus-data/

Siddiqi, A., A. Blair, and A. Parnia. (2020, April 16). A lack of data hides the unequal burden of COVID-19. *Toronto Star.* https://www.thestar.com/opinion/contributors/2020/04/16/a-lack-of-data-hides-the-unequal-burden-of-covid-19.html

Statistics Canada (2013). Immigration and Ethnocultural Diversity in Canada. *Statistics Canada* 99-010-X2011001. https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-010-x/99-010-x2011001-eng.pdf

Thebault, R., Ba Tran, A., & Williams, V. (2020, April 7). The coronavirus is infecting and killing black Americans at an alarmingly high rate. *Washington Post*. https://www.washingtonpost.com/nation/2020/04/07/coronavirus-is-infecting-killing-black-americans-an-alarmingly-high-rate-post-analysis-shows/?arc404=true.

Veenstra, G., & Patterson, A. C. (2016). “Black-White health inequalities in Canada.” *Journal of Immigrant and Minority Health, 18,* 51-57

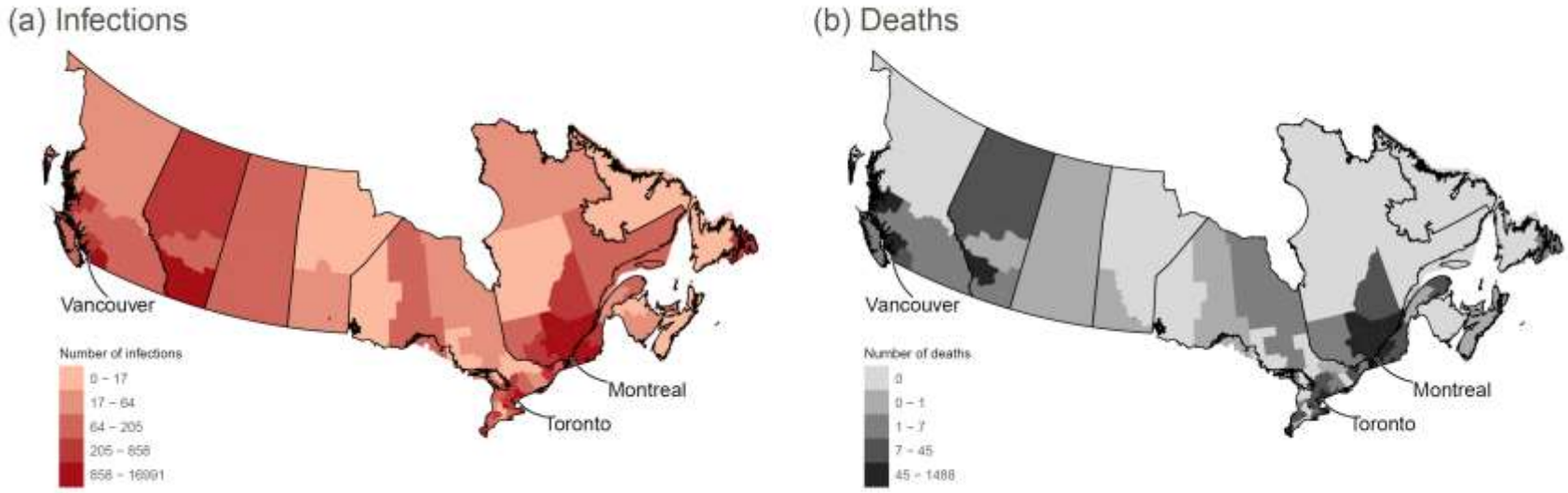
Yancy, C. (2020). COVID-19 and African Americans. *JAMA. Doi:* 10.1001/jama.2020.6548

FIGURES

Figure 1. Community Characteristics

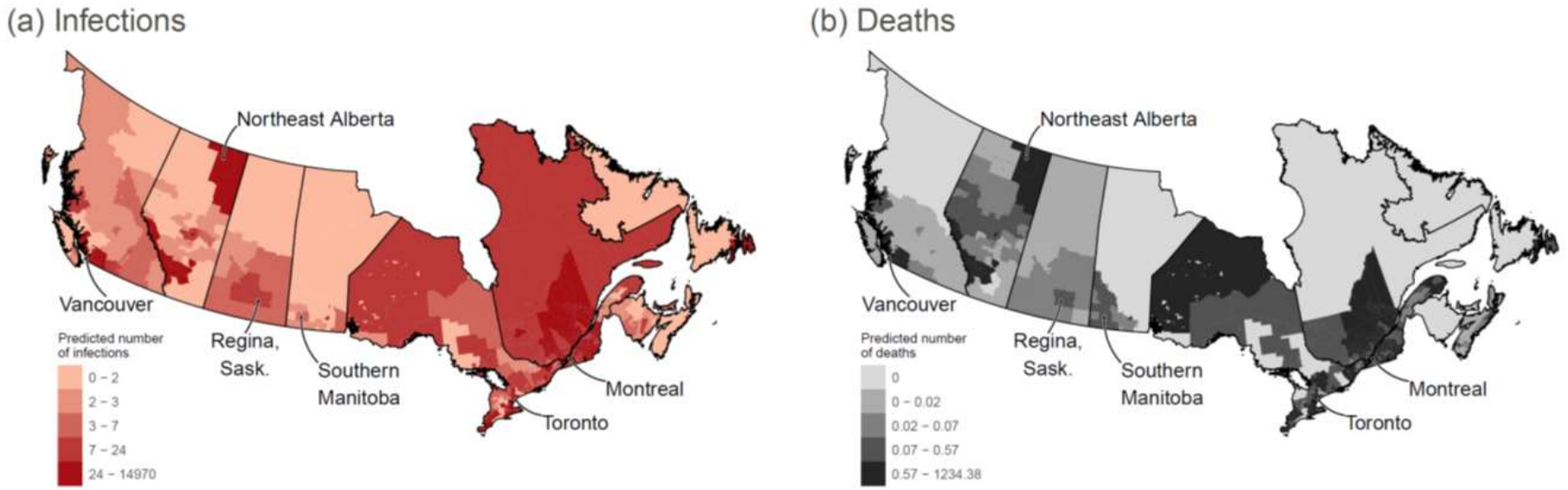
Source: Tabular data from Statistics Canada computed from 25% sample of the 2016 Census
Sample: 967 Pseudo-CSDs
Notes: 61,725 COVID-19 infections and 4,167 COVID-19 deaths in Canada on 5/5/2020

Figure 2. COVID-19 infection and death counts by health regions



Source: COVID-19 data released by the Public Health Agency in Canada
Sample: 82 health regions in Canada
Notes: 61,725 COVID-19 infections and 4,167 COVID-19 deaths in Canada on 5/5/2020

Figure 3. Predicted COVID-19 infection and death counts



Source: COVID-19 data merged with tabular census data
Sample: 967 Pseudo-CSDs
Notes: 61,725 COVID-19 infections and 4,167 COVID-19 deaths in Canada on 5/5/2020

TABLES

Table 1. Negative Binomial Regression Predicting Counts

	Infections				Deaths			
Characteristics of health regions	e^{β}	β		β/se	e^{β}	β		β/se
% Black	2.07	0.73	***	3.89	2.14	0.76	***	3.89
% Foreign-born	1.03	0.03	*	2.12	1.05	0.05	*	2.12
Median after-tax income	1.02	0.02		0.25	1.05	0.05		0.25
Population density (in 100s)	1.11	0.10		1.45	0.93	-0.07	*	-1.6
% 65 years and older	0.93	-0.07		-1.6	1.26	0.23		1.45
Intercept	5.86	1.77		0.52	0.01	-4.91		0.52

Sources: COVID-19 data released by Public Health Agency of Canada and tabular census data was compiled by Statistics Canada.
Sample: 82 health regions
Notes: *** $p<0.001$; **$p<0.01$; *$p<0.05$

ONLINE APPENDIX

Appendix tables

Table A1. Descriptive Analysis

	Mean	S.D.	Min	Max
B. Characteristics of health regions				
% Black	1.78	2.15	0.19	9.55
% Foreign-born	12.67	12.19	0.00	53.17
Median after-tax income	31,066	3,319	25,277	42,383
% 65 years and older	18.18	6.10	2.72	52.10
Population density (100s)	2.05	6.71	0.00	43.34
N		*82*		
B. Characteristics of pseudo-CSDs				
% Black	1.04	1.53	0.00	16.67
% Foreign-born	9.68	9.49	0.00	67.52
Median after-tax income	31,681	4,951	15,613	63,100
% 65 years and older	19.11	6.10	2.72	52.10
Population density (100s)	3.40	6.71	0.00	54.93
N		*967*		

Sources: COVID-19 data released by Public Health Agency of Canada and tabular census data was compiled by Statistics Canada.
Sample: 82 health regions (Panel A); 967 CSD/CDs (Panel B)
Notes: *** p<0.001; **p<0.01; *p<0.05

Table A2. Negative Binomial Regression Predicting Counts (Zero-order associations)

	Infections				Deaths			
Characteristics of health regions	e^{β}	β		β/se	e^{β}	β		β/se
% Black	1.72	0.54	***	5.37	1.71	0.54	***	4.56
% Foreign-born	1.08	0.08	***	5.47	1.09	0.09	***	4.96
Median after-tax income	0.94	-0.06		-1.22	0.88	-0.13	*	-2.01
Population density (100s)	1.11	0.11	**	3.24	1.15	0.14	**	2.65
% 65 years and older	0.80	-0.23	***	-3.80	0.82	-0.20	*	-2.51

Sources: COVID-19 data released by Public Health Agency of Canada and tabular census data was compiled by Statistics Canada.
Sample: 82 health regions
Notes: *** $p<0.001$; **$p<0.01$; *$p<0.05$

Table A3. Negative Binomial Regression Predicting Counts (Standardized Coefficients)

	Infection				Deaths			
Characteristics of health regions	e^{β}	β		β/se	e^{β}	β		β/se
% Black	4.78	1.56	***	3.89	5.11	1.63	***	3.14
% Foreign-born	1.46	0.38	*	2.12	1.88	0.63	*	2.65
Median after-tax income	1.07	0.06		0.25	1.18	0.17		0.55
% 65 years and older	0.61	-0.49		-1.6	2.65	0.98	*	-1.14
Population density	1.54	0.43		1.45	0.63	-0.46		2.44
Intercept	319.15	5.77		39.45	14.93	2.70		14.08

Sources: COVID-19 data released by Public Health Agency of Canada and tabular census data was compiled by Statistics Canada.
Sample: 82 health regions
Notes: *** p<0.001; **p<0.01; *p<0.05