The Conundrum of COVID-19: The Case of Toronto

A Bayesian framework for investigating the differential impact on lives and likelihoods across the city.

The unprecedented emergence of COVID-19 upended our lives to a great extent, resulting in pressing health crises and economic fallouts on a global scale. There have been, by and large, appreciable variations in the course of the COVID-19 outbreak across countries and territories. To come within the scope of this study, we combined resident-level COVID-19 fatalities data in Toronto, the 2016 census demographics data and the community council data to quantify the differentials in the extent to which residents in four Toronto districts are susceptible to COVID-19 during the early phase of the pandemic. We modelled the probability that individuals in different age groups would pass on after contracting COVID-19 in 2020 by the Hierarchical Logit Model. The findings are that these probabilities differ across four districts, which is attributable to varied capacities and unequal access to hospitals among neighbourhoods to a certain degree. Our aim is to provide data-based guidance for further research in public health policies to reduce health inequities

Authors

Quynh P. Vu
(quynh.vu@mail.utoronto.ca)
Prof. Samantha-Jo Caetano

Affiliations

Department of Statistical Sciences,
University of Toronto
Statistical Society of Canada (SSC)

01 Background

COVID-19 is a contagious disease that results from the novel strain of the SARS-CoV-2 virus. The first interhuman transmission case was confirmed in Toronto (Ontario) on January 25, 2020, shortly followed by widespread disruption to businesses, education, and essential health services across Canada. The Canadian government gave priority to the procurement and distribution of effective COVID vaccines to lift restrictions on our regular physical activities and stabilize outbreak incidence above all things. However, three years of the pandemic accentuated the inadequacy of hospital capacity and shortage of healthcare workers to serve individuals with medical needs, evidenced by an increase in physician office visits by 27% for the first twelve months of the pandemic. It poses a challenge to Canada's self-sufficiency and the ability to mitigate future outbreaks and curtail fatalities. It is noteworthy that "the concentration of poverty in particular neighbourhoods" resulting from income polarization in Toronto renders residents living in low-income neighbourhoods subject to poorer healthcare infrastructure and, therefore, more vulnerable to the pandemic than their counterparts. Estimating the individual mortality risk of COVID-19 (i.e. how likely a person is to die when infected with the disease) can help nip future outbreaks in the bud by guiding policymakers to implement public services and develop infrastructure that addresses those who are most in need. In this study, we examine the extent to which how mortality risk from COVID-19 varies across four districts in Toronto (Etobicoke York, North York, Toronto and East York, and Scarborough) after controlling for district-level population size and demographic age structure, sex at birth, and healthcare services.

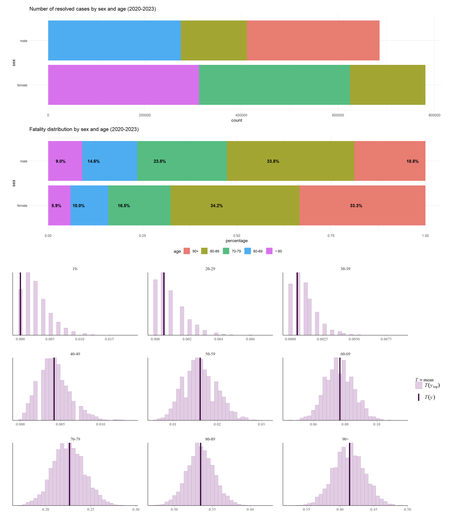
02 Data & EDA

We sourced the COVID-19 mortalities and Canadian 2016 neighbourhood-level census demographics from Open Data Toronto. This individual snapshot of mortality risk includes the patient's age and gender, neighbourhoods characterized by the Forward Sortation Area (FSA) code, date reported, whether hospitalized or not, and outcome (resolved or fatal). We then combined these with the community council data scraped from the City of Toronto website and the list of hospitals in Toronto by neighbourhoods scraped from Wikipedia. Our EDA indicates that, when offsetting the district population size, residents of Toronto and East York district are the least vulnerable to the pandemic.

Table 1: Pandemic data summary (as of April 10, 2023)

	North York	Toronto	Scarborough	Etobicoke	Pooled Statistics
Population (2016)	205887015	384239155	229218285	130116910	949461365
Prop. of 70+ residents (%)	15.05	13.661	15.997	15.49	14.734
No. of hospitals	9	20	4	2	35
No. of neighborhoods	41	74	23	22	160
No. of infected residents	70618	123027	60994	43266	297905
No. of fatalities	1170	1489	1117	733	4509
No. of residents per hospital	22876335	19211958	57304571	65058455	27127468
Avg. mortality risk (%)	1.657	1.21	1.831	1.694	1.514

The data sets are now available at https://github.com/pquynhvu/Bayes/tree/main/SSC2023/data.



03 Methodology

Let i = 1, 2, ..., N index infected individuals, j = 1, 2, ..., 9 index age groups, and k = 1, 2, 3, 4 index districts. For age group j and district k, the model can be written as

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}\,(\pi_i) \\
\eta_i &= \beta_0 + \beta_1\ \text{hospitalized}_i + \beta_2\ \text{sex}_i + \alpha^{\text{age}}_{j[i]} + \alpha^{\text{district}}_{k[i]}\ \text{admission rate}_{k[j]} \\
\pi_i &= \text{logit}^{-1}\,(\eta_i) = \tfrac{e^{\eta_i}}{1+e^{\eta_i}} \\
\beta_0, \beta_1, \beta_2 &\sim \text{N}\,(0,1) \\
\alpha^{\text{age}}_1 &\sim \text{N}\,(0,1)\,,\ \text{for } j = 2, \ldots, 9 \\
\alpha^{\text{age}}_j &\sim \text{N}\,(\alpha^{\text{age}}_{j-1}, \sigma^2_{\text{age}})\,,\ \text{for } j = 2, \ldots, 9 \\
\alpha^{\text{district}}_k &\sim \text{N}\,(0, \sigma^2_{\text{district}})\,,\ \text{for } k = 1, \ldots, 4 \\
\sigma^2_{\text{age}} &\sim \text{N}^+\,(0,1) \\
\sigma^2_{\text{district}} &\sim \text{N}^+\,(0,1)
\end{aligned}
$$

We fitted the proposed model in Stan using five chains and 1500 iterations (750 warm-ups). We placed weakly informative priors on all parameters and modelled the age coefficients as a first-order random walk.

While the model performs relatively similarly across four districts, it underestimates the risk among populations less than 40 years old, and except for the higher-than-expected age effect on the 19- population, this effect appears to increase as individual ages. Moreover, the model suggests that residents living in Scarborough are subject to the highest odds of dying, consistent with the average latent mortality risks provided in Table 1, while residents of the North York community are the least vulnerable. More importantly, as for a biological male and a biological female in the same cluster, the male would experience high odds of dying if infected with COVID-19. We also found that adding the hospital admission rate to the model improved the fit significantly.

04 Results & Discussion

We verified the effects of age and district of residence on one's probability of dying when infected with COVID-19. The result is no surprise that the senior population is subject to higher odds of dying, which varies for residents in the same age group but living in different districts. In particular, given two Toronto residents in the same age group, the one living in a less developed district (e.g., Scarborough) is more vulnerable to the pandemic than the one in a more developed area (e.g., North York). However, there is still room for improvement in future analysis. Firstly, regarding our data sets, the census data is from 2016, with an under-report rate of roughly 5%, which does not reflect the current demographic structure in Toronto. Moreover, we should not count on a single criterion to assess healthcare quality; measures such as the number of board-certified physicians, the number of ICU beds, or the ratio of providers to patients would give more meaningful results. Also, it would have been more interesting to add a temporal component to the model and compare how these probabilities change compared to the baseline probabilities of dying before vaccine rollouts.