2023 SSC ANNUAL MEETING

Quynh Vu

May 26, 2023

Abstract

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 communities 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 communities, 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.

Contents

1. Introduction

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 (Urrutia et al. 2021), 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 (Chang 2022). 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" (Hulchanski et al. 2010) 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 communities in Toronto (Etobicoke York, North York, Toronto and East York, and Scarborough) after controlling for community-level population size and demographic age structure, sex at birth, and healthcare services.

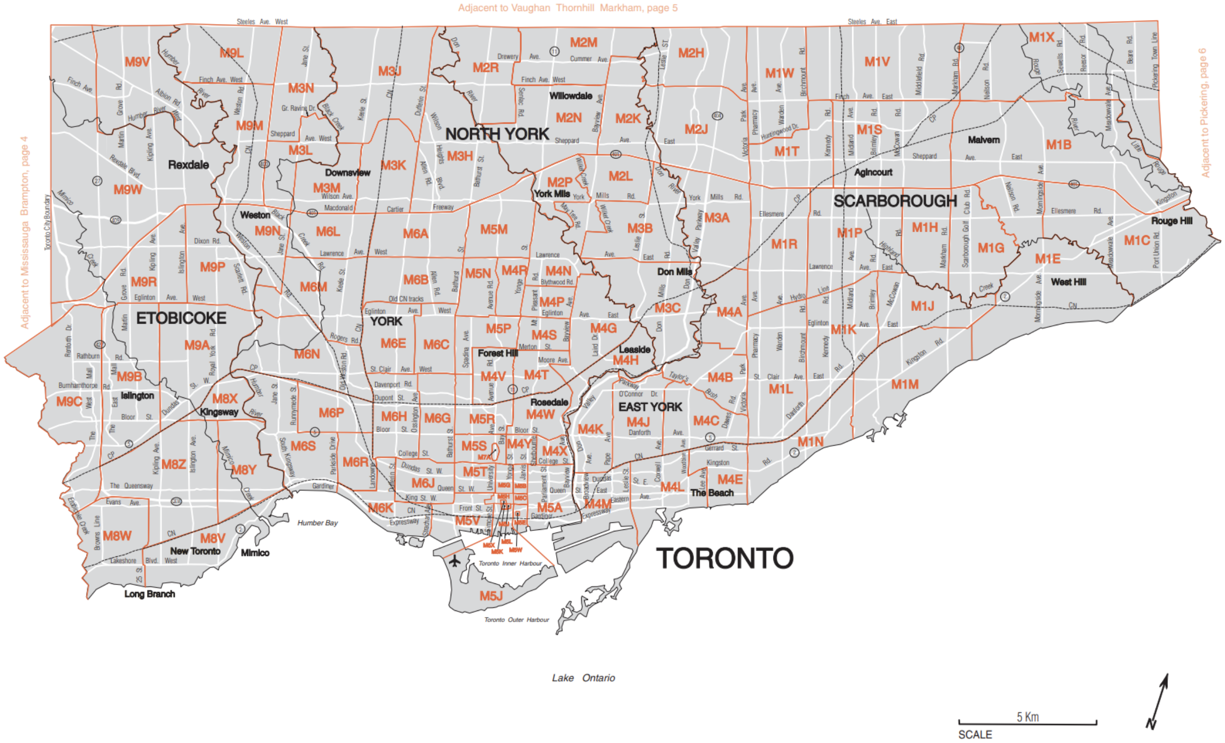


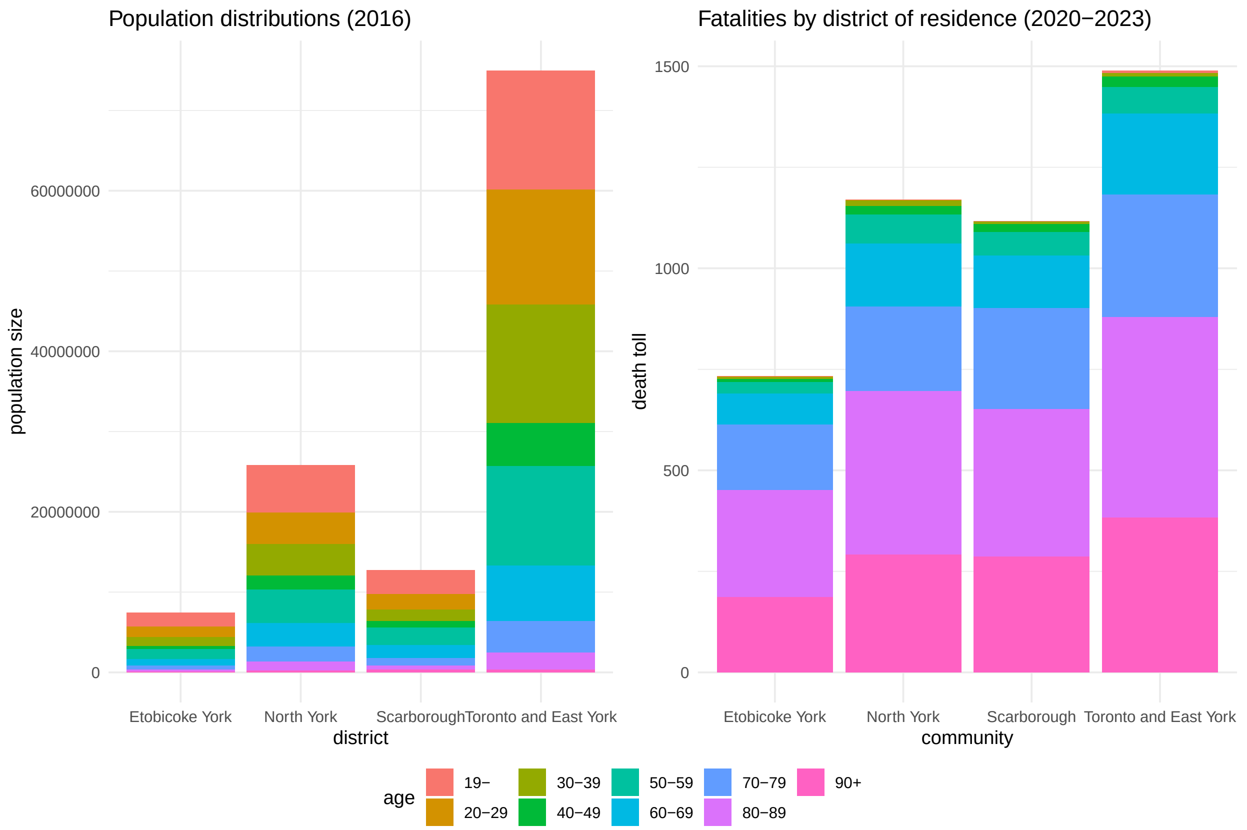
Figure 1: Toronto FSA map (found here)

2. Data

a) Data Sources and Cleaning

We combined 3 data sets for the analysis as follows:

- For the COVID-19 mortalities in Toronto data set, we sourced data from Open data Toronto, which is now hosted on GitHub (access here). 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 want to note that a "fatal" outcome implies any case that has died and the medical cause of death is related to COVID-19, while a "resolved" outcome means any case that has either recovered or died but the medical cause of death is unrelated to COVID-19.
- For the Canadian 2016 neighbourhood-level census demographics, we also sourced data from Open data Toronto, which is also hosted on GitHub (access here). The data set was aggregated from the total population in the 2016 Census by Statistics Canada and Toronto's 160 neighbourhood planning areas by the City of Toronto. For the 2016 census, the undercoverage rate published by Statistics Canada was 4.32% (Bérard-Chagnon and Parent 2021), which is the missed rate in the census due to travelling, refusal to participate, the growing number of immigrants and non-permanent residents in Canada, etc.
- For the community council data, we scraped data from the City of Toronto website using the `rvest` package to categorize Toronto's neighbourhoods into the designated communities. We also scraped the list of hospitals in Toronto by neighbourhoods from Wikipedia.



We obtained resident-level fatalities data in Toronto from 2020 to 2023. To study how the mortality risks from COVID-19 varied across communities in Toronto before pharmaceutical public health interventions were introduced, we primarily used resident-level fatalities data before December 15, 2020, when Ontario started its first phase of vaccine rollouts. We first matched the resident-level fatalities data with the list of hospitals in Toronto and categorized neighbourhoods into four designated communities by the FSA code. We created a numeric variable for the number of hospitals in each community (`num_of_hospitals`). We then aggregated this data set with the population data by matching the neighbourhood of residence and age group of each individual. To merge these two data sets, we corrected some differences in neighbourhoods' recorded names (e.g. *Danforth East York* to *Danforth-East York* or *Briar Hill-Belgravia* to *Briar Hill - Belgravia*). We created two numeric variables, the total population in each community (`pop_district`), and the average number of residents that a hospital in a particular neighbourhood should be able to serve at its maximum capacity, assuming that residents do not visit a hospital outside their community of residence (`pop_per_hospital`). Also, since there is mounting evidence that a biological male suffers more severe COVID-19 symptoms and has a higher mortality risk compared to a biological female (Scully et al. 2020), we only included individuals with clearly identified sex at birth in the analysis (i.e. we considered a transwoman a biological male and a transman a biological female and excluded patients who declared themselves to be non-binary, transgender, and other) to investigate these sex differences in the immune response against coronavirus within the Toronto population. The "cleaned" data set available to run analyses has no missing values.

The statistics summary table below provides an intuitive sense of the differences in outbreak settings across four districts and reveals relationship between age structure, medical facilities, and COVID-19 fatalities.

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

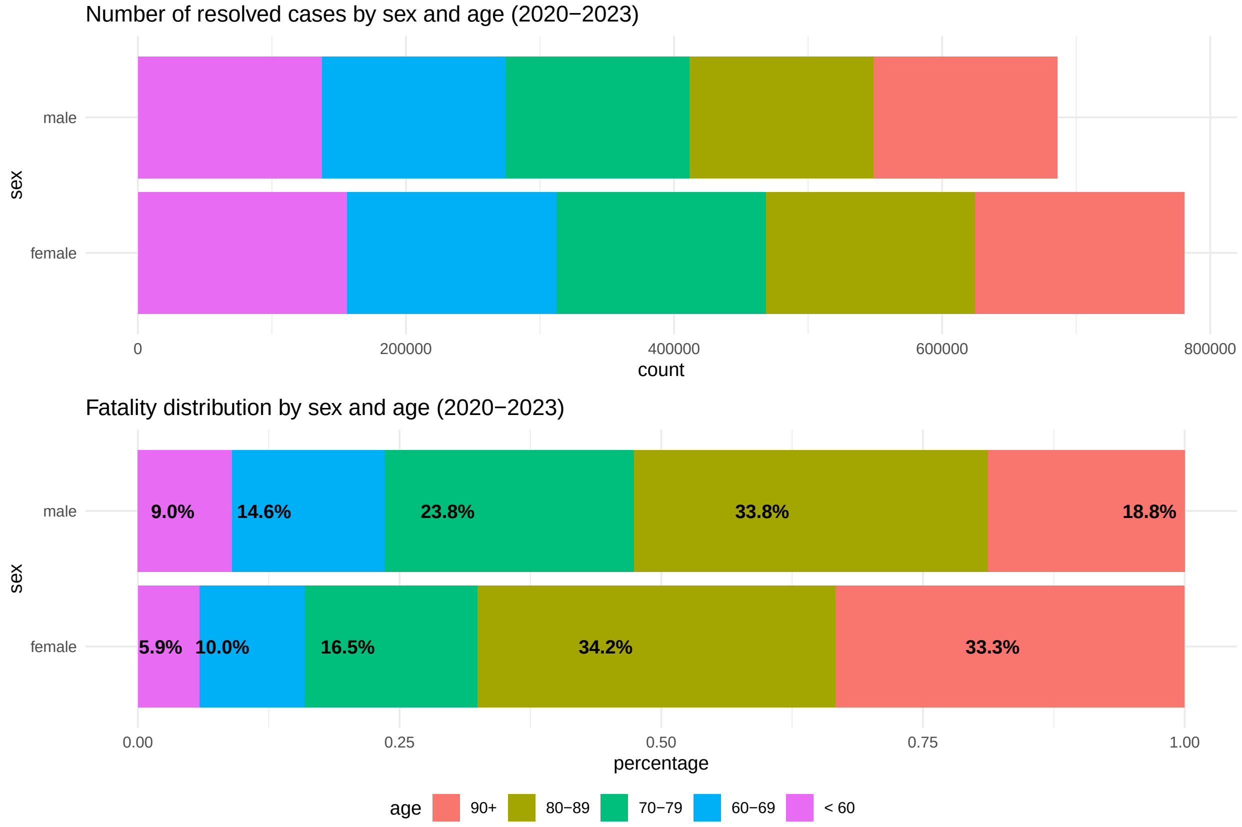
	North York	Toronto	Scarborough	Etobicoke	Pooled Statistics
Population (2016)	205887015	384239155	229218285	130116910	949461365
Prop. of 70+ residents (%)	15.05	13.661	15.997	15.49	14.734
No. of hospitals	9	20	4	2	35
No. of neighborhoods	41	74	23	22	160
No. of infected residents	70618	123027	60994	43266	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

b) Exploratory data analysis (EDA)

According to the 2016 census, Toronto residents above 70 account for roughly 12% of Toronto's population, but overall about 80% of COVID-19-infected patients who died in Toronto between 2020 and 2023 are above 70. As can be seen, Toronto and East York is the most populous community and experienced the highest death toll from COVID-19 in Toronto. However, if we are to offset the population size of each community by calculating the average of the latent death risk over COVID-19 contracted individuals in a community (Olsen et al. 2020),

$$\mu_k = \frac{\sum_{i=1}^{N_k} 1_{\text{fatal = 1, resolved = 0}}}{N_k}$$

where N_k is the total number of infected residents in district k, then residents in the Toronto and East York community are the least vulnerable to the pandemic, which is also reflected by its lowest proportion of infected residents who died due to COVID-19. Based on the data available, we hypothesize that the mortality risk varies across communities as a consequence of medical facilities (i.e., hospital capacity) since the low average number of residents per hospital is followed by a low latent death risk in the community and vice versa. Taking these into consideration, we studied individual mortality risk of COVID-19 prior to vaccine rollouts across four communities in Toronto. The primary dependent variable of interest was the probability of dying if infected with COVID-19 (1 = fatal, 0 = resolved) at the early stage of the pandemic as we sought to examine how the difference in available resources at hospitals across four communities had on mortality risks. The figure below shows that fewer males recovered from COVID than females. We also witnessed that most COVID-19 mortalities were in the 80+ population.

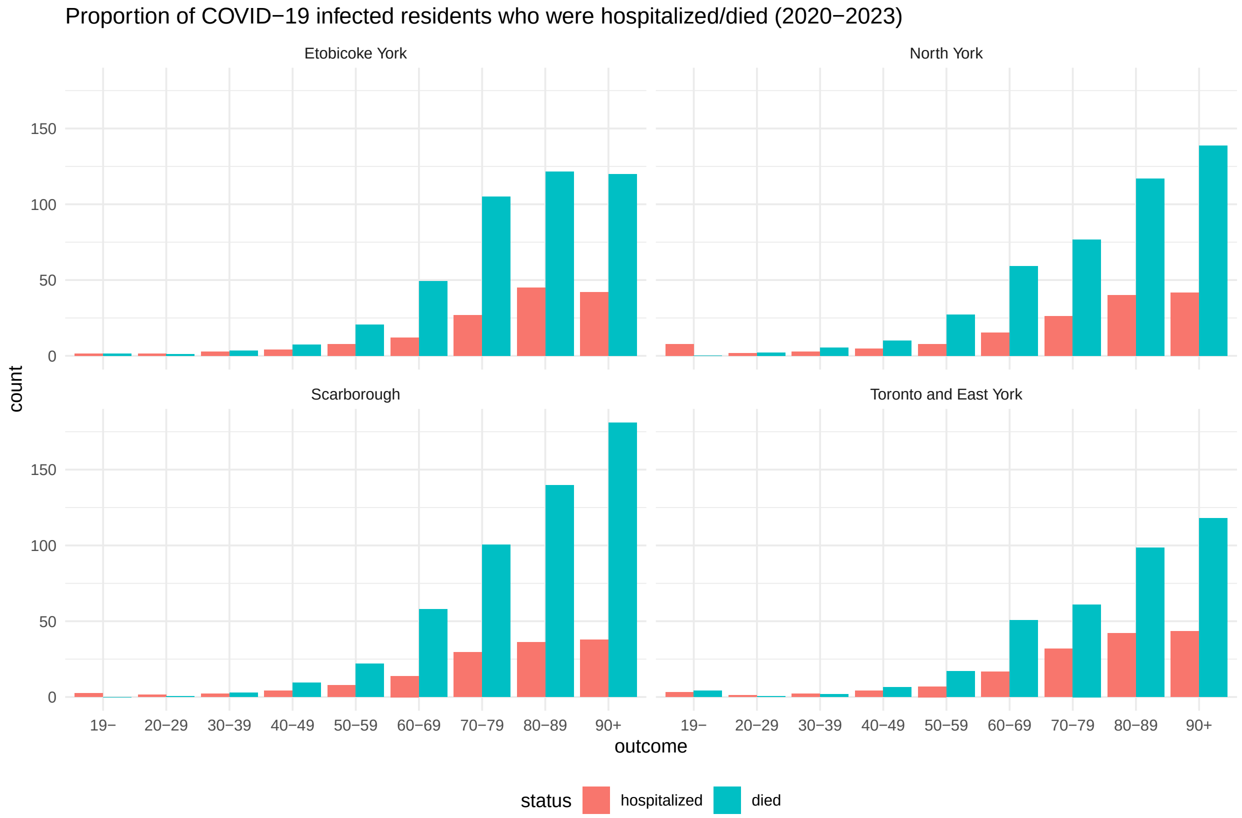


As mentioned, we used hospital capacity as a measure of medical facilities and healthcare quality measures, which, we acknowledge, should not be judged by this criterion alone. But more on this later. For now, our primary independent variables are summarized in the table below.

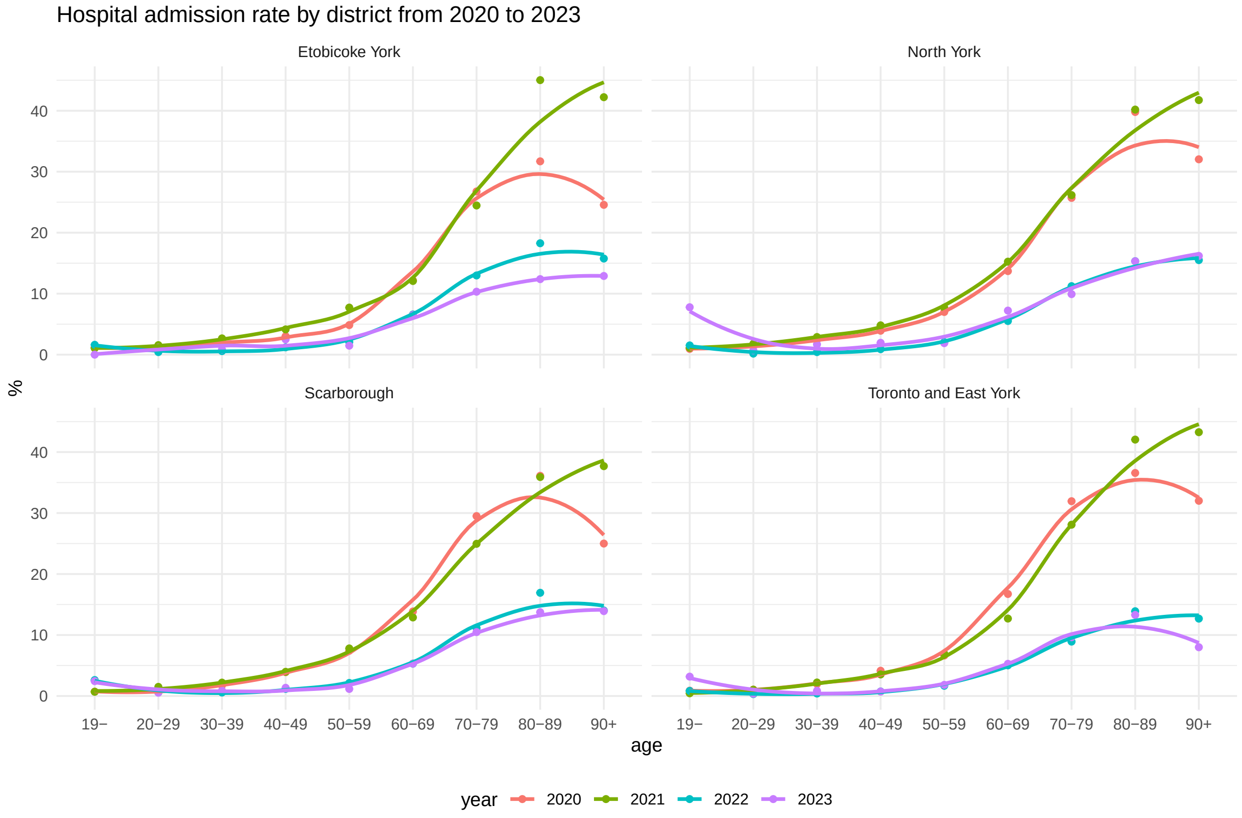
Table 2: Independent variables

Variables	(Coded) Value
whether the patient was hospitalized	1 = yes; 0 = no
sex	1 = female; 0 = male
age group	1 = 19-; 2 = 20-29; 3 = 30-39; 4 = 40-49; 5 = 50-59; 6 = 60-69 7 = 70-79; 8 = 80-89; 9 = 90+
district	1 = North York; 2 = Toronto and East York; 3 = Scarborough 4 = Etobicoke York
hospital admission rate	(numeric)

To justify our use of the indicator variable `hospitalized` and the hospital admission rate as independent variables, the figure below indicates that, given the varied number of hospitals across four communities, most COVID-19-infected patients admitted to hospitals were above 70 years old and the 90+ population experienced the highest death toll, and more importantly, across nine age groups, there are communities with higher hospital admission rate (i.e. higher hospitalized proportion) than others.



Proportion of COVID−19 infected residents who were hospitalized/died (2020−2023)



Hospital admission rate by district from 2020 to 2023

3. Methods

a) The model

As expressed previously, the probability of dying of residents in Toronto is not independent of each other given the predictor variables, i.e., individuals in the same "cluster" (same age group and same community) appeared to be more similar to each other than they were to individual in different clusters. This implies that biases in the standard errors arose since the independence assumption within clusters was no longer valid. With that in mind, we constructed a model with a hierarchical structure to compensate for these biases and proposed a Bayesian Hierarchical Logit Model to analyze the extent to which biological sex, age, and community of residence are related to the mortality risk of COVID-19.

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

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

where $y_i = 1$ if the i^{th} patient died due to COVID-19 and 0 otherwise, hospitalized_i and sex_i take binary values, and $\text{admission rate}_{k[j]}$ is the rate a hospital in community k admits patients in age group j.

Likelihood:

$$\text{likelihood} = \prod_{i=1}^{N} \pi_i^{y_i}(1-\pi_i)^{1-y_i} = \prod_{i=1}^{N} \big(\frac{e^{\eta_i}}{1+e^{\eta_i}}\big)^{y_i}\big(1-\frac{e^{\eta_i}}{1+e^{\eta_i}}\big)^{1-y_i}$$

Posterior distribution via Bayes Theorem:

$$| \text{ posterior} = \text{likelihood} \times \Big[\textstyle\prod_{l=1}^{3} \frac{1}{\sqrt{2\pi}} e^{\frac{-\beta_l^2}{2}}\Big] \times \Big[\textstyle\prod_{k=1}^{4} \frac{1}{\sqrt{2\pi}\sigma_{district}} e^{\frac{-(\alpha_k^{district})^2}{2\sigma^2_{district}}}\Big] \times \Big[\frac{1}{\sqrt{2\pi}} e^{\frac{-(\alpha_1^{age})^2}{2}}\Big]$$

$$| \times \Big[\textstyle\prod_{k=2}^{9} \frac{1}{\sqrt{2\pi}\sigma_{age}} e^{\frac{-(\alpha_k^{age}-\alpha_{k-1}^{age})^2}{2\sigma^2_{age}}}\Big] \times \Big[\frac{1}{\sqrt{2\pi}} e^{\frac{-\sigma^2_{age}}{2}}\Big] \times \Big[\frac{1}{\sqrt{2\pi}} e^{\frac{-\sigma^2_{district}}{2}}\Big]$$

b) Model Fitting and Validation Strategies

We fitted the proposed model in Stan using five chains and 1000 iterations (500 warm-up iterations) to the first three-month data on file (4468 data points between January 23, 2020, and April 23, 2020) due to limited technical resources. As the research question is to estimate the probability of dying across nine age groups and four communities in Toronto, we think it is justified to use the data at the very beginning of the pandemic to demonstrate the difference between clusters before pharmaceutical interventions.

To fit this model, we controlled for cluster-level attributes in our data by the α_j^{age} and $\alpha_k^{district}$ parameters. We modelled α^{age} as a first-order random walk with variance σ^2_{age}. We also explicitly parameterized variation across four communities by the $\sigma^2_{district}$ parameter. We placed weakly informative priors on all parameters as we did not want the priors to contribute strongly to the posterior distribution so we could make objective inferences about the parameter.

At first glance, the R-hat and n_{eff} for all parameters are less than 1.05 and greater than 100 (see **Table 3**), respectively, suggesting the between- and within-chain estimates agree. To further validate the model, we examined the trace plots of all parameters and carried out some posterior predictive checks (PPCs, LOO-CV, and test statistics) in the next section to compare the observed data to the data generated from our model.

Bérard-Chagnon, Julien, and Marie-Noëlle Parent. 2021. "MS Windows NT Kernel Description." 2021. https://www150.statcan.gc.ca/n1/pub/91f0015m/91f0015m2020003-eng.htm.

Chang, Bernard P. 2022. "The Health Care Workforce Under Stress—Clinician Heal Thyself." *JAMA Network Open* 5 (1): e2143167–67.

Hulchanski, J David et al. 2010. "The Three Cities Within Toronto." *Toronto: Cities Centre*.

Olsen, Wendy, Manasi Bera, Amaresh Dubey, Jihye Kim, Arkadiusz Wiśniowski, and Purva Yadav. 2020. "Hierarchical Modelling of COVID-19 Death Risk in India in the Early Phase of the Pandemic." *The European Journal of Development Research* 32 (5): 1476–1503.

Scully, Eileen P, Jenna Haverfield, Rebecca L Ursin, Cara Tannenbaum, and Sabra L Klein. 2020. "Considering How Biological Sex Impacts Immune Responses and COVID-19 Outcomes." *Nature Reviews Immunology* 20 (7): 442–47.

Urrutia, Deborah, Elisa Manetti, Megan Williamson, and Emeline Lequy. 2021. "Overview of Canada's Answer to the COVID-19 Pandemic's First Wave (January–April 2020)." *International Journal of Environmental Research and Public Health* 18 (13): 7131.