UNIVERSITY OF BRITISH COLUMBIA - OKANAGAN CAMPUS

Faculty of Science - Department of Computer Science, Mathematics, Physics, and Statistics

&

UNIVERSITY OF TORONTO - ST. GEORGE CAMPUS

Faculty of Science - Department of Statistical Sciences

COVID-19 Data for Long Term Care Homes in Canada

Written by

Shahrukh Alvi

Kaitlyn Hobbs

Dr. Ngan Lyle

Maggie Ma

October 2020

1 Executive Summary

In Canada, the impact of COVID-19 has been most tragic in long-term care (LTC) homes, with an estimated 50% of COVID-19 related deaths occurring in LTC homes[1]. For this project our group has leveraged multiple publicly available data sources to examine the geospatial, operational and structural factors contributing to the spread COVID-19 in Canadian LTC homes.

This project builds upon the Ontario COVID-19 analysis that was initiated by the Master of Data Science students, Kaitlyn Hobbs, Ngan Lyle, Sofia Bahmutsky and Shreeram Murali as a part of their Capstone project for Statistics Canada. For the original Capstone project, a database of LTC homes including home characteristics in Ontario was created. Preliminary analysis of this database had uncovered a number of factors that are significantly associated with whether or not a home had a COVID-19 outbreak. Therefore, in collaboration with Patrick Brown's research team from the University of Toronto, this study expands upon the initial analysis to include geospatial assessment of the spread of COVID-19 among LTC homes in Ontario. In addition, data were collected for LTC homes in British Columbia, Alberta, Manitoba, Saskatchewan, and Quebec.

To date, data collection and curation are complete for Ontario and British Columbia enabling modeling with LASSO regression, multivariate logistic regression and spatial regression. Surprisingly, basic data about the COVID-19 outbreak among LTC homes in Alberta, Manitoba, Saskatchewan, and Quebec were not readily available from reliable sources, and no meaningful analyses of these data were possible.

Results from our analysis of data from Ontario and British Columbia confirm that the size of a LTC home is the most important risk factor for a COVID-19 outbreak. In Ontario, municipal home type is a significant protective factor, while in British Columbia health authority ownership was not significantly associated with outbreak status. Results from the spatial analysis shows a spatial correlation with respect to the likelihood of an outbreak for homes in Southern Ontario but not with respect to the number of deaths.

Ongoing and future work includes creation of an R package with cleaned datasets for Ontario and British Columbia as well as preparation of a report for submission to the Canadian Journal of Statistics special issue "COVID-19: Statisticians in Action" by the end of the year. In this way we hope to show that open data can be used to provide insights, and to encourage participation in an open data environment.

2 Collaborations

Our team has worked closely with Professor Patrick Brown and Maggie Ma from the University of Toronto's Department of Statistical Sciences, who conducted the spatial modeling of LTC homes in Ontario. We also met with interested parties at Statistics Canada to update them on our progress and explore opportunities for collaboration. In fact, one member of our team, Kaitlyn Hobbs, has been offered a position as a data scientist at Statistics Canada which has strengthened our collaboration. A summary of our team's major meetings is shown in the Appendix, Table 1.

Data on LTC homes in British Columbia were supplied by Rob Cowan-Douglas of Seniors Advocate BC, an advocacy organization that is affiliated with the Government of British Columbia. Mr. Cowan-Douglas also provided valuable insights into some of the limitations of the data.

3 Methodology

3.1 Data retrieval

All web scraping and data extraction was performed in python.

British Columbia
Data regarding LTC facility structure, history of inspections, infractions, and complaints, as well as quality indicators were obtained from Seniors Advocate BC. COVID-19 outbreak information, including the cumulative number of cases and deaths through to August 21, 2020, was obtained by email from British Columbia's Senior Public Affairs Officer, Stephen May.

Alberta
Data was scraped from Alberta Health's Continuing Care Branch's list of facilities[2] and the Government of Alberta's web page on COVID-19 cases[3]. Alberta Health's 2016 list of facilities was used as the number of facilities were in agreement with the number of homes (177) reported by the Canadian Institute for Health Information[4]. Unfortunately, available data are missing several important variables, including the number of beds in each facility, which had already been observed to be a significant predictor in Ontario. Moreover, outbreaks are declared over once four weeks have passed without additional cases, and facilities with resolved outbreaks are removed from the government web page. Information regarding whether or not an LTC had reported a COVID-19 outbreak was collected on September 10, 2020.

Manitoba

Data was extracted or scraped from the following sources: a personal care home list accessible on the internet from manitobanurses.ca website[5]; an email with covid outbreak data from Carla Loeppky, the Director and Lead Epidemiologist of Epidemiology and Surveillance at Manitoba Health which received on September 29, 2020; and a bed numbers list from the Long-Term and & Continuing Care Association of Manitoba (LTCAM) website[6]. The geocoder.ca website was used to get latitude and longitude values for the homes. Unfortunately, the data are missing several variables necessary for effective analysis. A response from a government contact stated that the "Manitoba Government was not willing to provide that level of data at this time".

Saskatchewan
Data was scraped from the Government of Saskatchewan's webpage on personal care homes[7] and the Open Database of Healthcare Facilities (ODHF) by Statistics Canada[8]. Geocoder.ca was used to get latitude and longitude values for homes where latitude and longitude values were not available from the ODHF. Unfortunately, an official source for COVID-19 outbreak data could be found.

Quebec
No reliable list of long term care homes in the province could be found. In addition, a cumulative list of homes that have experienced an outbreak was not available. The Government of Quebec website lists only homes with an active outbreak and like in Alberta, homes with a resolved outbreak are removed from the list[9].

3.2 Data cleaning and merging

All data were processed using python. For each province, several datasets were merged to create a database of LTC home data and COVID-19 outbreak data. Flowcharts of data merging and processing for Ontario and British Columbia are shown in the Appendix, in Figures 1 and 2.

3.2.1 Merging Statistics Canada's Open Database of Health Facilities

The latitude and longitude coordinates for most of the LTC homes were taken from Statistics Canada's Open Database of Health Facilities (ODHF), which was merged with provincial data on a 'cleaned name' variable. Mismatched facility names were manually resolved. After merging with ODHF, any remaining missing coordinates were manually entered using an online geocoder.

3.2.2 Cleaning the British Columbia Data

The British Columbia data was partitioned into two datasets: (1) a complete dataset comprised of the LTC home names, locations, health authority, ownership type (non-for-profit, for-profit, municipal), direct care hours, quality related variables and

outbreak status and (2) a dataset containing only variables without missing values. In this paper, the aforementioned datasets will be referred to as the “complete” and the “selected” datasets, respectively.

4 Results and Analysis

All statistical analyses were carried out using R. Due to insufficient data from other provinces, only the Ontario dataset and the selected British Columbia dataset were analyzed.

4.1 LTC Homes Analysis

Data suitable for analysis was available for Ontario and British Columbia. LASSO regression was applied to the datasets. We then looked at the order in which variables were added to the model as the lambda increased in order to guide variable selection for further analysis. The covariates ultimately chosen for analysis for each province are shown below in Table 1.

Table 1. Variables chosen for further analysis.

Ontario	**British Columbia**
Number of beds	Number of beds
Home type (Municipal, For-profit or Non-profit)	Ownership (Health authority, For-profit or Non-profit)
Presence of short stay beds	Regulation (Hospital act, Community care and assisted living act or both)
Family council	Family council
Residents council	Residents council
Accreditation	Accreditation
Number of inspections in the last 2 years	Presence of shared rooms
Number of inspections with orders in the last 2 years	Number of disease related infractions in the last year

4.1.1 Logistic Regression

The relationship between the variables chosen for analysis and the binary outcome of outbreak versus no outbreak was modeled using logistic regression. In Ontario, the likelihood of an outbreak is significantly associated with the increased number of beds (Beta = 0.014, $P < .001$) and the number of inspections that resulted in one or more orders in the last 2 years (Beta = 1.45, $P = .04$). Conversely, municipal home type (Beta = -0.75, $P = .006$), the presence of short stay beds (Beta = -0.46, $P = 0.04$) and the number of inspections that did not result in any orders (Beta = -956, $P = 0.04$) was protective. In British Columbia, the likelihood of an outbreak is significantly associated with an increased number of beds (Beta = 0.017, $P < .001$)) and the presence of a family council was protective (Beta = -1.21, $P = .02$).

4.1.2 Spatial Regression

Two generalized linear geostatistical models were fit to measure the spatial effect of deaths and outbreaks in Southern Ontario nursing homes. Both models included number of beds, nursing home types, and total number of inspections and complaints as predictors, and each accounted for individual nursing home effects. Latitude and longitude information were converted to euclidean distance using the Pmisc package in R and transformed to a projection suitable for the Toronto region (epsg = 4326). A 100km buffer was also applied to both models in order to reduce edge effects.

The first model used the number of deaths as the outcome variable with a zero inflated censored poisson distribution. However, the spatial results produced were not statistically significant, meaning if two nursing homes were side by side geographically, they could have very different outcomes with respect to residents' deaths. The only statistically significant variable from the first model was the type of home; residents have a higher chance of dying from COVID in a for-profit home (baseline), than a not for profit (-0.86) or a municipal home (-1.64). The second model used a binomial distribution to assess whether a LTC home that had an outbreak is correlated with their physical proximity to other LTC homes. The result of this model is displayed below in Figure 1. A spatial effect is present and statistically significant. Some hot spots of COVID-19 cases in LTC homes include Ottawa, Windsor, London, Waterloo, Niagara, and the Greater Toronto Area.

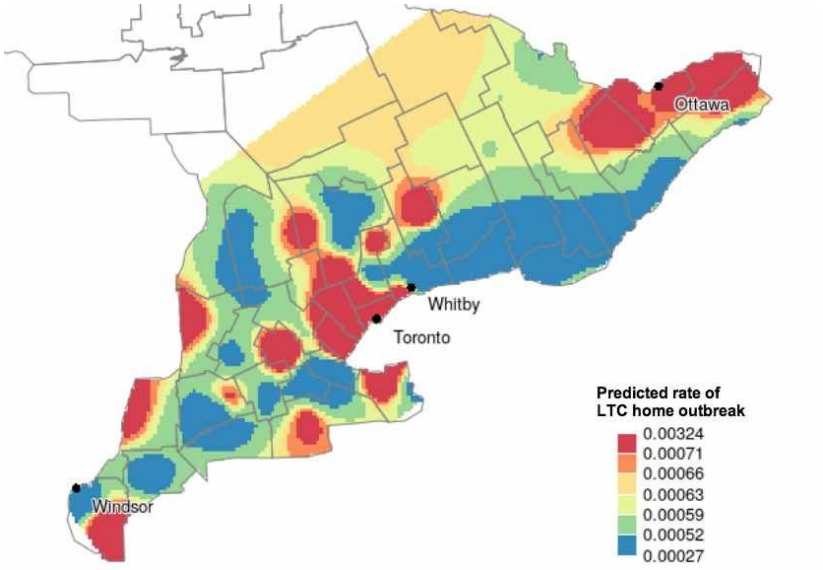

Figure 1. Visualization of the spatial logistic regression.

5 Challenges

5.1 Inaccessible or incomplete government data

There were several challenges to data collection. Firstly, outside of Ontario, it was difficult to find reliable and accessible sources of information. For example, while Alberta Health has a listing of LTC homes from 2019, the data was accompanied with a warning that they may be incomplete or erroneous. Secondly, outbreak reporting was either absent (in Saskatchewan and Manitoba) or limited to current outbreaks (in Alberta and Quebec). For Quebec, an attempt was made to access the historical version of the website using the WayBack website, but this was not possible. Also, when governments were contacted for information by email or by phone, inquiries were met with either non-reponse (for Alberta and Quebec) or a negative response (for Manitoba). Even for British Columbia, while we were able to get good information about outbreaks, this information was emailed privately. Thirdly, inconsistencies in available variables and differences in how an “outbreak” is defined between provinces, make interprovincial comparisons difficult. For example, in British Columbia a LTC home is reported as having an outbreak with detection of a single case, while in Alberta a home is not classified as having an outbreak until there have been at least five cases.

In summary, complete data suitable for analysis were available only for Ontario and British Columbia. For the Manitoba dataset, predictors such as number of beds, longitude and latitude were compiled. However, a comprehensive list of proprietary (for-profit) and non-proprietary (non-profit) homes could not be found. For the

Saskatchewan dataset, we have information about the number of beds, the number of infractions, initial license date, monthly rates, single rooms, double rooms, latitude and longitude. The critical missing data for this province is COVID-19 outbreak status. For Quebec neither a list of all the homes in the province nor one with homes that have had an outbreak could be found. Therefore, no meaningful dataset could be constructed.

5.2 Data with incongruent or missing values

Among the provinces with adequate data for analysis, incongruent or missing values were particularly problematic for British Columbia. For example, the dataset includes data on the number of single-occupancy, double-occupancy and multi-bed rooms for each LTC home. However, some of the numbers are inaccurate, as they are self-reported. In our discussions with the data analytics manager at Seniors Advocate BC, it was suggested that these covariates be converted to binary variables indicating the presence or absence of each room type.

Missing values was another significant problem with the British Columbia data. A matrix of the missing values is shown below in Figure 2. Moreover, most of the missing values are not “missing at random”, meaning that analyses omitting those homes would result in bias. For example, complaints data are missing for all homes in Northern Health that are regulated under the Hospital Act. This affects 12 out of the 24 homes in Northern Health. Similarly, incident data are missing for all homes in Vancouver Island Health that are regulated under the Hospital Act, affecting 22 out of the 59 homes in that health authority. As a result, while the original dataset contains 37 potentially useful predictors, when variables with missing values are excluded, only 15 predictors are left.

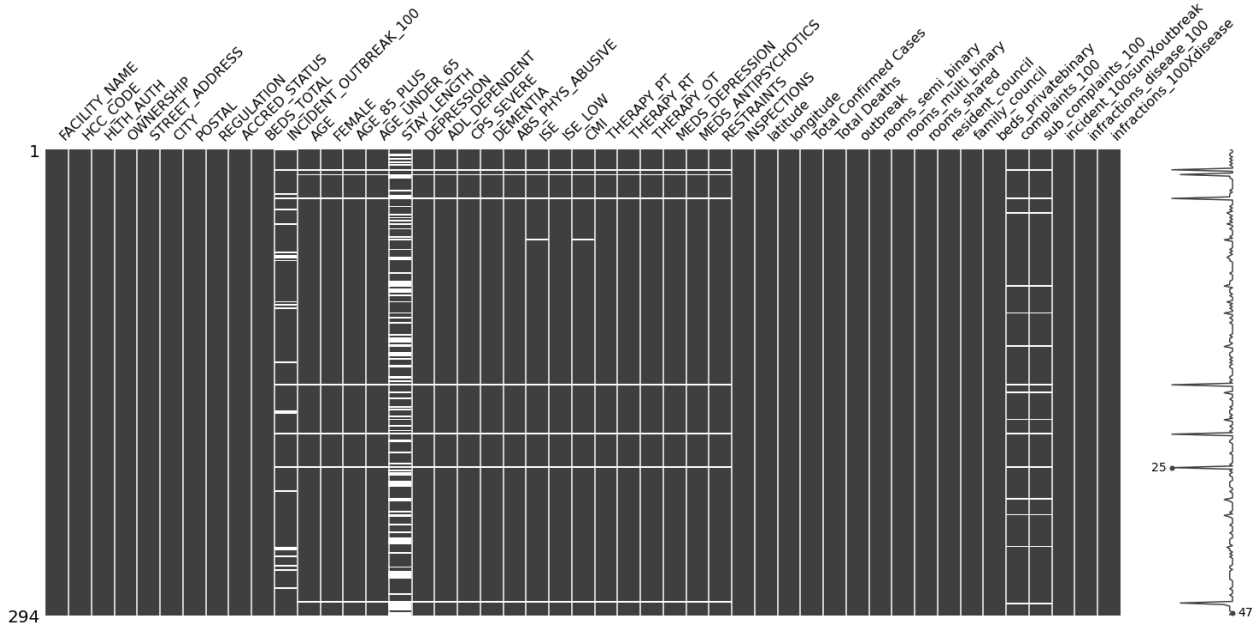

Figure 2. Missing values matrix generated for the British Columbia LTC homes dataset using the missingno package in python.

5.3 Inter-provincial data variability

While both British Columbia and Ontario publish data for LTC homes in their respective provinces, how the variables are defined can and often do vary by province. As such, inter-provincial comparisons are difficult and have to be explored with caution. For example, although data from Ontario suggests that municipally operated homes are protected against outbreaks, a similar effect for health authority operated homes in British Columbia was not observed. In this comparison, it is important to keep in mind that British Columbia does not have municipally operated LTC homes. Instead, homes may be operated under a health authority, which standardizes operations at a much higher level of aggregation than a municipality. Therefore, it may be that the level of government at which a home is operated impacts on a home's resilience to a COVID-19 outbreak.

6 Conclusions

In conclusion, our goal for this project was to create a national database of LTC homes. Over the last few months we have collected or have attempted to collect data for Ontario, British Columbia, Alberta, Manitoba, Saskatchewan and Quebec. To our surprise, basic data at the level of individual LTC homes, for example a list of homes that have had an outbreak, were either not accessible or difficult to access through official government sources, for all provinces except Ontario. This may explain why the

majority of peer-reviewed publications to date about the spread of COVID-19 among LTC homes in Canada, focuses on the experience in Ontario.

For the two provinces for which suitable data were available, Ontario and British Columbia, LASSO regression, logistic regression and spatial regression were applied for modeling. LASSO regression was used to guide variable selection for the other two methods. In Ontario, our logistic regression model showed that municipal home type, the presence of short stay beds and the number of inspections in the last 2 years were significantly associated with the likelihood of an outbreak. Interestingly, while the total number of inspections in the last 2 years is negatively associated with the likelihood of an outbreak, the number of infractions, or inspections that resulted in an order by the inspector, was positively associated with the likelihood of an outbreak. This suggests that, in Ontario, regular inspections may play a role in maintaining or improving standards and that non-compliances may identify homes in need of attention. In British Columbia, apart from the number of beds, logistic regression showed that the presence of a family council was negatively associated with the likelihood of outbreak. At this point, spatial regression has been performed for Ontario only. Initial results show that there is a spatial effect on the likelihood of a home experiencing an outbreak but not on the number of COVID-19 related deaths. Altogether, our analyses suggest that the size of a home is the most important risk factor for a COVID-19 outbreak, that location is important and that in Ontario municipal home type is protective.

7 Future Work

Given its timeliness and relevance, our group is committed to continuing to work on this project. Currently, our focus is on completing the analysis of the Ontario and British Columbia datasets. Priorities include, applying the spatial models to the British Columbia data as well as modeling case and death numbers along with temporal data for Ontario. Our goal is to submit a report for publication in the Canadian Journal of Statistics special issue “COVID-19: Statisticians in Action” by the end of the year.

In addition, we are working on creating an R package, CanadaCovid19, which will house our cleaned datasets. In the future, this package can be used by learners to demonstrate how basic statistics can provide insight into complex real world problems. In addition, interested researchers will be able to rerun our analyses as well as conduct their own analyses. Our hope is to stimulate discussion and collaboration in an open data environment, and to help advance the understanding of COVID-19.

8 References

1. Weeks, C., Mahoney, J., Stone, L. & Ha, T. T. Outbreaks at seniors' homes linked to almost half of COVID-19 deaths in Canada, Theresa Tam says. *The Globe and Mail* (2020).
2. List of publicly funded designated supportive living accommodations and long-term care facilities - Open Government. (2016).
3. Cases in Alberta. https://www.alberta.ca/covid-19-alberta-data.aspx.
4. Profile of Residents in Residential and Hospital-Based Continuing Care, 2018–2019.
5. Provincial PCH Bed Map list provided by Shared Health.pdf. https://manitobanurses.ca/system/files/Provincial%20PCH%20Bed%20Map%20list%20provided%20by%20Shared%20Health.pdf.
6. PCHBedNumbersOct2019_000.pdf. (2019).
7. Government of Saskatchewan. Personal Care Homes. http://personalcarehomes.saskatchewan.ca/PersonalCareHomes/Map.
8. Government of Canada, S. C. The Open Database of Healthcare Facilities. https://www.statcan.gc.ca/eng/lode/databases/odhf (2020).
9. Situation of the coronavirus (COVID-19) in Québec. https://www.quebec.ca/en/health/health-issues/a-z/2019-coronavirus/situation-coronavirus-in-quebec/.
10. Government of Ontario. How Ontario is responding to COVID-19. https://www.ontario.ca/page/how-ontario-is-responding-covid-19 (2020).
11. Search Selections for Long-Term Care homes. http://publicreporting.ltchomes.net/en-ca/Default.aspx.
12. Health Quality Ontario. Long-Term Care Home Performance in Ontario. https://www.hqontario.ca/System-Performance/Long-Term-Care-Home-Performance (2020).
13. Huber, N. British Columbia Long Term Care Facilities Quick Facts Directory 2019. (2019).

9 Appendix

Data (both raw and cleaned), selected scripts, presentations and other documentation can be found at https://github.com/nganlyle/cansii_covid_proj.

Table 1. Summary of Meetings

Date	Activity	Description of Activity	Documentation
July 20, 2020	UBCO team meeting	UBC Okanagan group members met with Professor Braun to establish goals for the project.	meeting_mins_07.20.2020.pdf
August 4, 2020	UBCO - UofT collaborative meeting	Collaborative meeting to review the initial results of the spatial analysis.	meeting_mins_08.04.2020.pdf
August 17, 2020	UBCO - UofT collaborative meeting	Collaborative meeting to review the updated results of the spatial analysis.	meeting_mins_08.17.2020.pdf
August 18, 2020	Statistics Canada collaborative meeting	Meeting with Bruno St. Aubin (Statistics Canada) and Professor Gabriel Wainer's lab (Advanced Realtime Simulation) to explore opportunities for collaboration.	Not available
August 21, 2020	UBCO team meeting	Meeting with Professor Braun to discuss progress on the collection and cleaning of the British Columbia LTC homes data.	meeting_mins_08.23.2020.pdf
September 10, 2020	Statistics Canada collaborative meeting	Meeting with members of the Data Science Division, PHAC modelling group at Statistics Canada to share results.	statscan_covid_presentation.pdf
September 24, 2020	UBCO - UofT collaborative meeting	Collaborative meeting to discuss progress on the collection of data from Quebec and Manitoba. Also, we discussed preliminary results from analysis of the British Columbia data.	meeting_mins_09.24.2020.pdf
October 19, 2020	UBCO - UofT collaborative meeting	Collaborative meeting to discuss finalizing the Ontario and British Columbia analysis and publication.	meeting_mins_10.19.2020.pdf

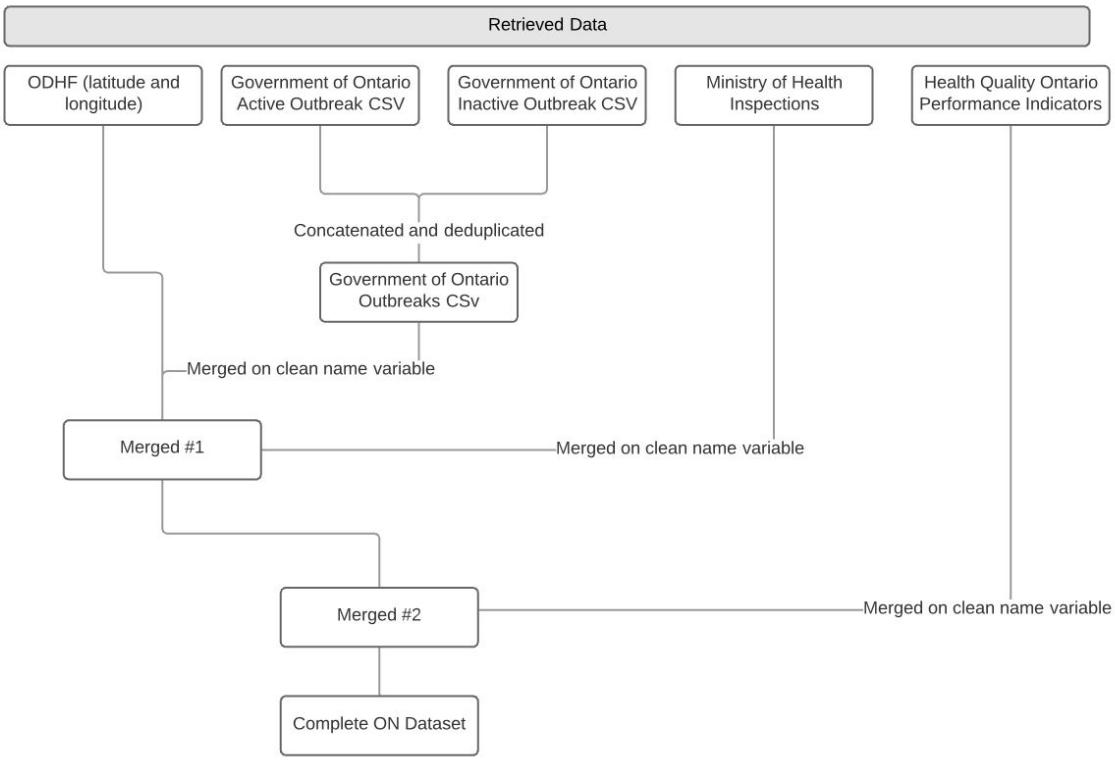

Figure 1. Flowchart of data processing for Ontario. Four datasets were merged: Statistics Canada's Open Database of Healthcare Facilities (Version 1.1)[8], Government of Ontario's listing of active and resolved long-term care COVID-19 outbreaks[10], Ministry of Health's Inspections[11], Health Quality Ontario's performance indicators for long-term care facilities[12].

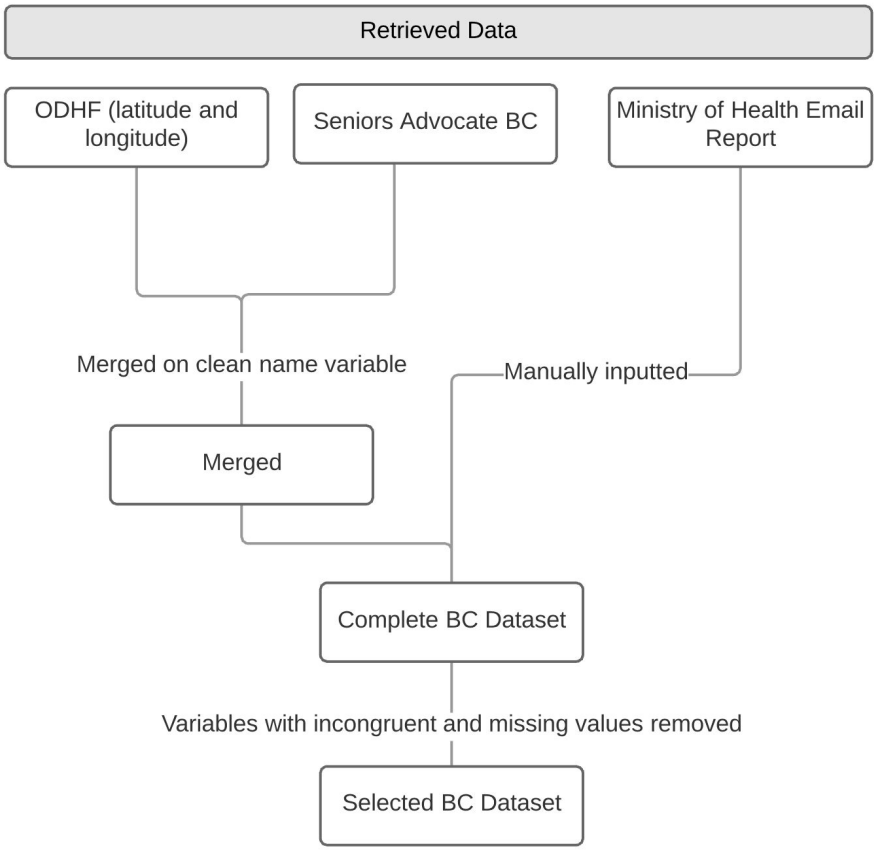

Figure 2. Flowchart of data processing for British Columbia. Three datasets were merged: Statistics Canada's Open Database of Healthcare Facilities (Version 1.1)[8], Seniors Advocate BC's LTC data[13], and The Ministry of Health's emailed report of LTC outbreaks in British Columbia.