Data Article

A longitudinal dataset of incidence and intervention policy impacts regarding the COVID-19 pandemic in Canadian provinces

Ammar Hassanzadeh Keshteli [a], Dana Allen [a], Afia Anjum [b],
Yashvi Patel [a], Aadhavya Sivakumaran [a], Siyang Tian [a], Fei Wang [b,c],
Hao Wang [d], Mark A. Lewis [a,d], Russell Greiner [b,c],
David S. Wishart [a,b,*]

[a] Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada
[b] Department of Computing Science, University of Alberta, Edmonton, AB T6G 2E8, Canada
[c] Alberta Machine Intelligence Institute, Edmonton, AB T5J 3B1, Canada
[d] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada

A R T I C L E   I N F O

A B S T R A C T

One year after identifying the first case of the 2019 coronavirus disease (COVID-19) in Canada, federal and provincial governments are still struggling to manage the pandemic. Provincial governments across Canada have experimented with widely varying policies in order to limit the burden of COVID-19. However, to date, the effectiveness of these policies has been difficult to ascertain. This is partly due to the lack of a publicly available, high-quality dataset on COVID-19 interventions and outcomes for Canada. The present paper provides a dataset containing important, Canadian-specific data that is known to affect COVID-19 outcomes, including sociodemographic, climatic, mobility and health system related information for all 10 Canadian provinces and their health regions. This dataset also includes longitudinal data on the daily number of COVID-19 cases, deaths, and the constantly changing intervention policies that have been implemented by each province in an attempt to control the pandemic.

* Corresponding author at: Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada.
E-mail address: dwishart@ualberta.ca (D.S. Wishart).

Specifications Table

Subject	Public Health and Health Policy
Specific subject area	Coronavirus disease 2019 (COVID-19) pandemic in Canada
Type of data	Table
How data were acquired	Data was collected from different online (primary) sources affiliated with Canadian federal or provincial governments, and provincial health authorities.
Data format	Raw, analyzed
Parameters for data collection	Data on socio-demographic, climatic, mobility (via Google cell phone records), daily numbers of new cases and deaths due to COVID-19, as well as COVID-19 intervention policies implemented by the 10 provincial governments in Canada from the beginning of February 2020 to the end of February 2021 were collected.
Description of data collection	Publicly available Canadian federal and provincial sources were used to extract data on socio-demographic and health-related characteristics, holidays, daily weather-related data, daily mobility data (based on cell phone records), daily number of new confirmed cases of COVID-19 and the corresponding number of deaths, hospitalizations, and ICU admissions in the 10 Canadian provinces and their affiliated health regions from February 1, 2020 to February 28, 2021. We also collected and summarized intervention policies introduced by each province to manage the pandemic from March 1. 2020 to January 13, 2021.
Data source location	Canadian provinces of Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland and Labrador, Nova Scotia, Ontario, Prince Edward Island, Quebec, and Saskatchewan. The primary data source locations for all provincial data and policy intervention data used to assemble the reported dataset are provided in Supplementary Table 1.
Data accessibility	Direct URL to data: https://doi.org/10.17632/hjfp9c5gdr.1

Value of the Data

- Almost a year after the start of COVID-19 pandemic in Canada, over 1.1 million Canadians have been diagnosed with the disease and more than 24,000 people have died [1]. This dataset combines comprehensive longitudinal data regarding COVID-19 incidence, interventions and intervention impacts across all 10 Canadian provinces and their affiliated health regions using reliable provincial and national resources. To the best of our knowledge, there is no other publicly available Canadian dataset of this kind with such rich longitudinal information.

- This dataset includes comprehensive sociodemographic, health-related, and mobility data that are known to be related to the burden of COVID-19 (e.g., incidence, mortality, etc.) at both the provincial and health-region levels. These data will enable researchers in different fields, such as public health, biostatistics, and epidemiology, to predict the trajectory of COVID-19 in different Canadian provinces using a variety of modelling techniques.

- In this dataset, we have categorized and carefully scored different intervention policies that were introduced by provincial governments across Canada. These data will provide a unique opportunity for researchers and policy makers to evaluate the effectiveness of different sets of policies and compare their impacts between provinces. These data could help pandemic modellers and policy makers design, develop, and introduce more appropriate policies to mitigate the COVID-19 pandemic, not only Canada but potentially in other countries.

1. Data Description

In this article, we present a dataset that provides daily information related to the COVID-19 pandemic in each of the 10 Canadian provinces and their health regions from February 1, 2020 to February 28, 2021. The dataset contains 10 manually generated XLSX format files: one file per province that includes 68 variables. Specifically, the data curation team manually collected, checked and entered the raw provincial and health region data onto pre-formatted Excel spreadsheets. Each of these files contains 8 sheets consisting of specific data types described in more detail below. Note that Sheets 1 and 2 are static – i.e., the data does not change from day to day. Sheets 3–7 present daily quantities.

Sheet 1 presents socio-demographic and health-related data for each province and its affiliated health regions are presented. The demographic data includes the following features: total population, percentage of different age groups (0–14, 15–39, 40–59, 60–79, and ≥80 years), percentage of females, percentage of diabetic patients, percentage of smokers, percentage of population with post-secondary education, median household income, number of meat packaging plants (facilities), number of intensive care unit (ICU) beds, number of ventilators, area (square kilometre (sq. km)), population density, number of private dwellings, and housing density. Most of this data is static and specific to the most recent year for which data has been publicly released (often 2019 or 2020).

Sheet 2 lists the distances (km) between the main airport in each province and main cities in each health region.

Sheet 3 presents weather data and includes daily mean temperature (°C) and total precipitation (mm) data from February 1, 2020 to February 28, 2021 [2].

Sheet 4 presents daily COVID-19-related data and includes day of week, weekday (presented as 1) or holiday (presented as 0), daily number of COVID-19 cases, number of tests administered, current hospital admissions, current ICU admissions, and daily COVID-19 deaths in 10 Canadian provinces and their health regions from February 1, 2020 to February 28, 2021. Fig. 1 shows the daily number of COVID-cases per 100,000 population in each of the 10 Canadian provinces from February 1, 2020 to February 28, 2021.

Sheet 5 presents the daily values of changes in Google mobility-related data [3] from a baseline level (The baseline day is the median value from the 5–week period Jan 3–Feb 6, 2020). This data includes changes related to retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential mobilities. As an example, Fig. 2 shows the changes in mobilities related to retail and recreation centers and changes in mobilities related to workplaces in the province of Alberta.
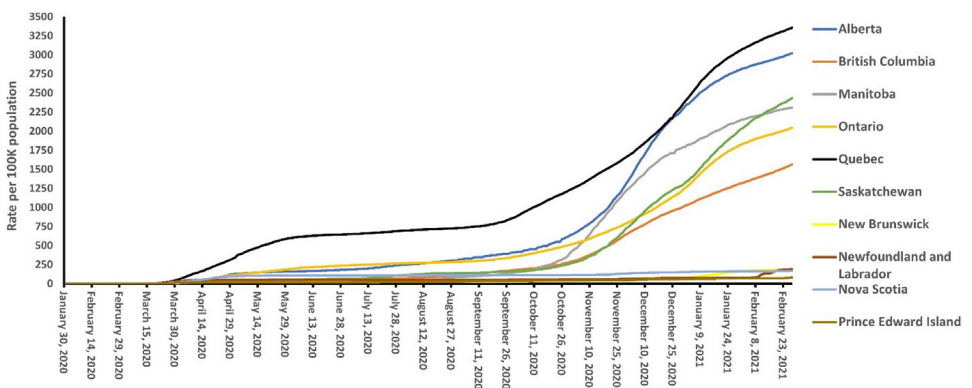

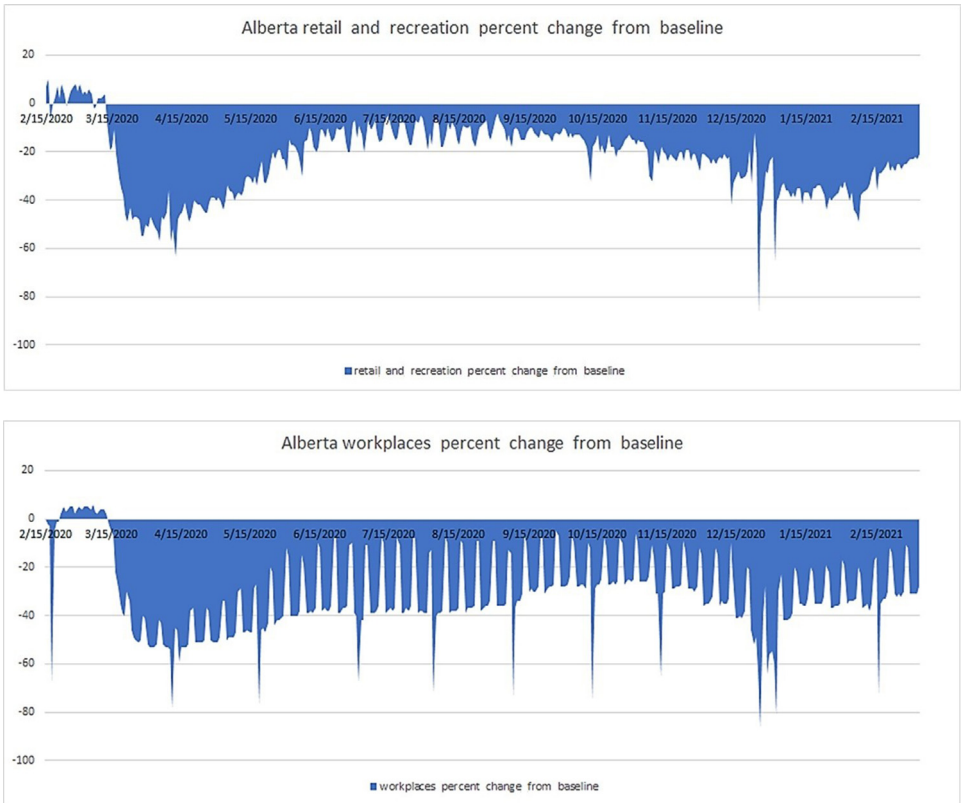**Fig. 1.** Cumulative rate of COVID-19 cases in Canadian provinces.

Fig. 2. Percentage of changes in mobilities related to retail/recreation centers and workplaces from baseline in the province of Alberta during the one-year course of COVID-19 pandemic. A baseline day represents a normal value for that day of the week. The baseline day is the median value from the 5-week period from Jan 3, 2020 to Feb 6, 2020 (as instructed by Google).

Sheet 6 presents the daily changes in stringency scores of COVID-19-related intervention policies across all 10 Canadian provinces from March 1, 2020 to January 13, 2021. As an example, Fig. 3 presents the daily changes in stringency scores of restrictions related to restaurants and international air travel in the Canadian provinces of Alberta and Newfoundland.

Sheet 7 presents the list of sources that were used to gather data.

2. Experimental Design, Materials and Methods

We used several online sources released or maintained by the Canadian federal or provincial governments as well as a number of highly regarded Canadian digital media outlets (which used their contacts within the provincial governments) to extract COVID-19-related data from the 10 Canadian provinces and their health regions from the beginning of February 2020 to the end of February 2021. In each spreadsheet file, the source(s) used to obtain the relevant information for each of the different data features is presented in sheet 8.

We used the following list of provinces and their affiliated health regions:

- Alberta (Calgary, Edmonton, Central, North, and South),
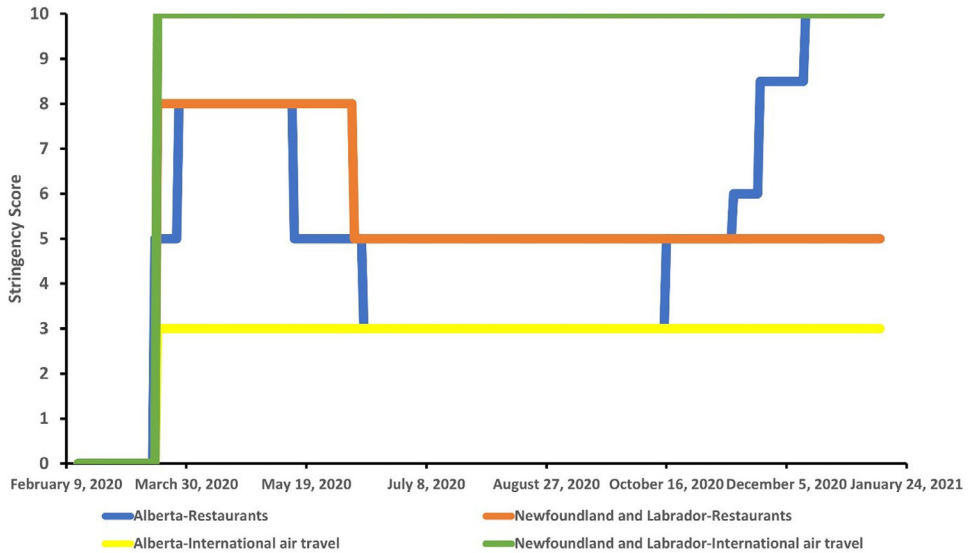- British Columbia (Fraser, Interior, Island, Northern, and Vancouver Costal),

Fig. 3. Daily changes in stringency scores of restrictions related to restaurants and international air travels in the Canadian provinces of Alberta and Newfoundland and Labrador.

- Manitoba (Interlake-Eastern, Northern, Prairie Mountain Health, Southern Health-Santé Sud, and Winnipeg)
- New Brunswick,
- Newfoundland and Labrador,
- Nova Scotia,
- Ontario (Toronto, Ottawa, Northern, Eastern, Central East, Central West, and South West),
- Prince Edward Island,
- Quebec (Nunavik Nord Du Quebec/Northern Quebec, Cote Du North/North Coast, Saquenay-Lac Saint Jean, Abitibi- Temiscamingue, Outaouais, Laurentides, Lanaudière, Mauricie - Center-du-Québec, Capitale-Nationale, Montréal, Laval, Montérégie, Estrie, Chaudière-Appalaches, Bas-Saint-Laurent, Gaspésie-Îles-de-la-Madeleine, and Terres-Cries-de-la-Baie-James)
- Saskatchewan (Far North, North, Central, Saskatoon, South, and Regina).

In the Atlantic provinces (i.e., New Brunswick, Newfoundland and Labrador, Nova Scotia, and Prince Edward Island), we only collected data at the provincial level due to the relatively small populations and/or small land areas of these provinces. Furthermore, we did not collect data for the Canadian territories due to the small population and low number of daily COVID-19 cases.

The sociodemographic data presented in Sheet 1, were mainly collected at the provincial level as we could not find reliable information on certain static features (e.g., number of ventilators or ICU beds) or temporal features (e.g., daily number of deaths or hospitalizations, intervention policies) at the health region level.

For the data collected on weather [2] (Sheet 4) and Google mobility [3] (Sheet 6), we used information from the most populous city in each health region as a representative of the daily weather or mobility status of that health region.

In the present study, we used COVID-19 Intervention Scan-Data Tables developed by the Canadian Institute for Health Information (CIHI) [4] to capture changes in COVID-19-related policies as they were introduced by provincial governments across Canada. This data provides information on selected interventions (e.g., policy decisions) intended to stop the spread of COVID-19 and improve health outcomes by different jurisdictions across Canada. Our team separated these intervention policies into 11 different groups and subgroups and scored each group from 0 to

10 (0: no restriction, 10: total restriction). These groups (and associated subgroups) are as follows: (1) Dine out services (subgroups: restaurants/cafe/lounges, bars/casinos/bingo halls), (2) Non essential businesses (subgroups: shopping malls/retail stores, industries/manufactures, financial/corporate workplaces), (3) Education (subgroups: universities/post secondary institutes, schools), (4) Child homes/childcare services, (5) Health services (subgroups: hospital visitation, long-term/continuing care visitation, non-essential hospital services, non-essential outpatient/clinic services), (6) Gathering size (subgroups: indoor gatherings, outdoor gatherings), (7) Traveling restriction (subgroups: domestic Travel, international air travel, international travel to US by bus or car), (8) Public services (subgroups: recreation, non-essential services), (9) Religious congregation restriction (subgroups: indoor gatherings, outdoor gatherings), (10) Testing strategies, and (11) declaration of a public health state of emergency. The most recent version of CIHI dataset contains interventions announced as of January 13, 2021 which is summarized in our dataset. Our dataset will be updated once the next version of CIHI COVID-19 Intervention Scan-Data Tables is released.

Ethics Statement

All data presented in this dataset is anonymized and aggregated at the health region or provincial level.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

CRediT Author Statement

Ammar Hassanzadeh Keshteli: Data curation, Methodology, Writing – review & editing; **Dana Allen:** Data curation, Methodology, Writing – review & editing; **Afia Anjum:** Data curation, Methodology, Writing – review & editing; **Yashvi Patel:** Data curation, Methodology, Writing – review & editing; **Aadhavya Sivakumaran:** Data curation, Methodology, Writing – review & editing; **Siyang Tian:** Data curation, Methodology, Writing – review & editing; **Fei Wang:** Data curation, Methodology, Writing – review & editing; **Hao Wang:** Writing – review & editing; **Mark A. Lewis:** Conceptualization, Writing – review & editing; **Russell Greiner:** Conceptualization, Writing – review & editing; **David S. Wishart:** Conceptualization, Writing – review & editing.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.107381.

References

[1] Government of Canada. Coronavirus disease 2019 (COVID-19): Epidemiology update. https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html?stat=num&measure=total&map=pt#a2. Accessed May 12, 2021.

[2] Meteorological service of Canada, environment and climate change Canada. Weather information. https://weather.gc.ca/mainmenu/weather_menu_e.html. Accessed May 12, 2021.

[3] Google. Community mobility reports. https://www.google.com/covid19/mobility. Accessed May 12, 2021.

[4] Canadian institute for health information. COVID-19 Intervention Timeline in Canada. https://www.cihi.ca/en/covid-19-intervention-timeline-in-canada. Accessed May 12, 2021.