Catalogue no. 13260002

User Guide and Data Dictionary for Preliminary COVID-19 Data

Release date: April 9, 2021
Updated: August 13, 2021

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Table of contents

1. Background 4

2. Objectives of the Preliminary dataset on confirmed cases of COVID-19 4

3. Coverage of the Preliminary dataset on confirmed cases of COVID-19 4

4. Content of the Preliminary dataset on confirmed cases of COVID-19[1] 5

5. Limitations 6

6. Data quality concerns 6

7. References 6

Appendix I - Data dictionary, Concept, Variable Name, Section on the Form, Description and Universe 7

Appendix II - Data dictionary, Notes and Limitations 8

Appendix III - Data dictionary, Source, Format and Answer Categories 9

Appendix IV – Information on the Week Variables, for the Year 2020 10

Appendix V - Information on the Week Variables, for the Year 2021 11

Appendix VI - Information on the Episode Week Group Indicator 12

User Guide and Data Dictionary for Preliminary COVID-19 Data

1. Background

COVID-19 is the disease caused by SARS-CoV2, a novel coronavirus that has not been identified before the first cases were reported in Wuhan, China, in December 2019. The virus has now spread to almost all countries around the world. The first confirmed cases in Canada appeared in January 2020.

There is still a lot that is unknown about the virus and limited data available for researchers to study it. In Canada, the 10 provinces and 3 territories are providing the Public Health Agency of Canada (PHAC) information on COVID-19 cases on a daily basis. In collaboration with PHAC, Statistics Canada (StatCan) contributes to make these preliminary data available to the research community and to all Canadians.

2. Objectives of the Preliminary dataset on confirmed cases of COVID-19

PHAC and StatCan have been working closely together to be able to provide preliminary data received by the provinces and territories (PTs) to researchers.

The Preliminary dataset on confirmed cases of COVID-19 provides easy access to as much data as possible, by provincial regions, while respecting confidentiality of the individuals for which information on COVID-19 history is reported.

Given the COVID-19 pandemic is still progressing, the content of this dataset will be updated regularly, making it a unique and relevant product. Each iteration of the dataset will provide up-to-date case information reported by PTs.

This information was originally released in the Detailed preliminary information on confirmed cases of COVID-19 (Revised) table but due to the increasing number of cases, this dataset could no longer be supported in this format. This table was deleted on Thursday, December 10th, 2020. The information from this table is now available in a downloadable dataset: Preliminary dataset on confirmed cases of COVID-19, Public Health Agency of Canada" (13-26-0003).

3. Coverage of the Preliminary dataset on confirmed cases of COVID-19

The data published by StatCan contains cases for which detailed case information was submitted by the provincial or territorial public health authority to PHAC. The governments of Canada and the provinces and territories agreed on a common Case Report Form (CRF)1 to be used to report cases to PHAC.

These data may not match the total cases reporting done at the provincial and territorial levels, which are updated daily by each jurisdiction and compiled by PHAC. The discrepancy is due to factors such as delays in reporting, or variability in reporting cut-offs. Given the under coverage, these data are a subset of the total reported cases in Canada.

Routine updates on health outcome status are not made uniformly across Canada, and therefore the data may underestimate the number of hospitalizations, admissions to intensive care units, deaths and recoveries.

Throughout the pandemic, the CRF has been updated, which has impacted the data published by StatCan. For example information about symptoms was removed from the dataset as of March 2021 since the information was no longer collected on the CRF and the historical information was incomplete.

The data on this dataset is preliminary and subject to change as updated information is received from the provinces and territories.

4. Content of the Preliminary dataset on confirmed cases of COVID-19[1]

This dataset is a subset of the information that provinces and territories collect using the *Coronavirus Disease (COVID-19) Case Report Form*. The variables selected were those that were considered to be the most important while meeting a certain quality threshold. Also, some "derived variables" were computed by PHAC based on the information contained in the case report forms.

To minimize the risk of disclosure:

1. a few categories from the original questions collected on the form have been grouped together:
    - ▶ The provinces and territories have been grouped into the following regions:
        - British Columbia & Yukon
        - Alberta, Saskatchewan, Manitoba & the Northwest Territories
        - Ontario & Nunavut
        - Quebec
        - New Brunswick, Nova Scotia, Prince Edward Island & Newfoundland and Labrador
    - ▶ The age in years of individuals has been grouped into age groups:
        - 0-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+
2. a few categories from the original questions have been reclassified:
    - ▶ Occupation
        - Health care workers, school or daycare workers/attendees and residents of long-term care facilities have remained as their own categories.
            - Please note that healthcare workers include those with and without direct patient contact. Long-term care residents may include residents of senior's homes, assisted living facilities, and retirement communities, as well as nursing homes. Long-term care facilities may be privately run or under provincial authority.
        - Farm worker, laboratory worker and veterinary/animal worker have been grouped with 'other'.
        - Occupations with small numbers have been grouped with 'other'.
3. a few derived variables have been created:
    - ▶ Dates (Episode Date, Onset Date, Resolution Date)
        - All dates were converted to weeks, i.e., Episode week, Episode week group, Onset week, Resolution week.
        - All cases with an episode date and/or onset date prior to February 23rd, 2020 (the first day of the 8th week of 2020) were grouped with the cases in week 8. For those cases that also had a resolution date, the resolution week was also shifted forward the same amount as the episode and/or onset date.
        - If there are insufficient cases by week for a region to be able to release the episode week without compromising the confidentiality, the cases for a given period will be grouped in the week that had the most cases (i.e., not enough cases for the Atlantic region for weeks 23, 24 and 25, most cases were in week 23, all cases will be grouped in week 23). The Episode week group variable will indicate if a grouping was done and appendix V will indicate which weeks have been grouped together. For those cases that also had an onset week and/or resolution week, the dates were also shifted to keep the same duration.

Refer to the data dictionary for detailed information about each variable.

The information on this dataset is considered **preliminary**.

5. Limitations

This dataset includes cases who are confirmed according to the Canadian interim national case definition for COVID-19, that is: "A person with laboratory confirmation of infection with the virus that causes COVID-19 performed at a community, hospital or reference laboratory (NML or a provincial public health laboratory) running a validated assay."[2]

COVID-19 testing was initially performed for diagnostic purposes only (i.e., to confirm the diagnosis of suspected cases of COVID-19), and then, increasingly, for screening based on public health priorities (high-risk groups or contact tracing). None of these tests were conducted for research purposes, and the screening was not designed to be conducted in a probabilistic sample representative of the Canadian population.

The expansion of laboratory testing evolved over time following the epidemiology of the disease, i.e., the spread of the disease from China to other countries and the establishment of community transmission in Canada.

With increasing laboratory capacity, some provinces were able to screen people from targeted groups, e.g., residents and staff of long-term care facilities where cases have occurred, or contacts of cases identified in epidemiologic investigations. These expansions of testing did not occur simultaneously across provinces and territories. Additionally, testing capacity and prioritization continue to differ between provinces and territories, thus skewing any inter-jurisdictional comparison.

The factors listed above must be taken in consideration when interpreting data analysis. Examples of possible bias include:

- One could observe that the proportion of symptomatic cases decreased over time, which is simply because initially, only symptomatic individuals were eligible to be tested.
- Following outbreaks in long-term care facilities, some jurisdictions undertook mass screening in residents of these facilities, which may impact the age distribution of cases. Mass screening in specific segments of the population may lead to their over-representation in the confirmed case data, as general population mass screening has not occurred on a large scale.
- Any comparisons between provinces and territories using demographics or health outcomes may be biased by differences in testing criteria.

6. Data quality concerns

Routine updates on health outcome status are not made uniformly across Canada, therefore the data may underestimate the number of hospitalizations, admissions to intensive care units, deaths and recoveries.

There is a high proportion of missing values and some sections of the case report form were filled inconsistently.

Please note that variables may be recoded or changed based on several factors, including but not limited to: new information being reported for historical cases, updates to the case report form, revised reporting by provinces and territories, etc. StatCan and PHAC are working closely together to improve the quality of the file with the help of all provinces and territories. This will be reflected in each iteration of this dataset.

StatCan and PHAC are working closely together to improve the quality of the file with the help of all provinces and territories. This will be reflected in each iteration of this dataset.

7. References

1. CORONAVIRUS DISEASE (COVID-19) CASE REPORT FORM https://www.canada.ca/content/dam/phac-aspc/documents/services/diseases/2019-novel-coronavirus-infection/health-professionals/2019-nCoV-case-report-form-en.pdf.
2. Interim national case definition: Coronavirus disease (COVID-19) https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/health-professionals/national-case-definition.html.

Appendix I - Data dictionary, Concept, Variable Name, Section on the Form, Description and Universe

Concept	Variable Name	Section on the Form	Description	Universe
Case identifier number	COV_ID	Administrative Information	Unique identifier for each case	All cases
Region	COV_REG	Administrative Information	Province/Territory where the case resides, grouped by regions.	All cases
Episode week	COV_EW	Administrative Information	Week of the episode, derived using the earliest of the following dates: symptom onset date, specimen collection date, laboratory testing date, date reported to the province or territory, or date reported to PHAC.	All cases
Episode week group	COV_EWG	Administrative Information	Indicates when multiple episode weeks have been grouped together to protect confidentiality. Refer to Appendix VI.	All cases
Episode year	COV_EY	Administrative Information	Year of the episode, derived using the earliest of the following dates: symptom onset date, specimen collection date, laboratory testing date, date reported to the province or territory, or date reported to PHAC.	All cases
Gender	COV_GDR	Case Details	The gender of the case. Where available, gender data was used; when gender data was unavailable, sex data was used.	All cases
Age group	COV_AGR	Case Details	Age group corresponding to the age of the case	All cases
Occupation	COV_OCC	Case Details	Indicates the case's occupation	All cases
Asymptomatic	COV_ASM	Symptoms	Indicates if the case was symptom free at time of diagnosis.	All cases
Onset week of symptoms	COV_OW	Symptoms	Week of symptom(s) onset	All cases
Onset year of symptoms	COV_OY	Symptoms	Year of symptom(s) onset	All cases
Hospital status	COV_HSP	Clinical Course and Outcomes	Indicates if the case was hospitalized and if the case was admitted to the intensive care unit.	All cases
Resolved	COV_RSV	Clinical Course and Outcomes	Indicates if the case has resolved.	All cases
Resolution week	COV_RW	Clinical Course and Outcomes	Week reported resolved	Resolved=1
Resolution year	COV_RY	Clinical Course and Outcomes	Year reported resolved	Resolved=1
Death	COV_DTH	Clinical Course and Outcomes	Indicates if the case died due to COVID-19 which may be attributed when COVID-19 is the cause of death or is a contributing factor	All cases
Transmission	COV_TRM	Exposures	Location where exposure occurred	All cases

Appendix II - Data dictionary, Notes and Limitations

Concept	Note and Limitation
Case identifier number	Created randomly by Statistics Canada. The same case will have a different number every time that the file is released.
Region	To ensure confidentiality, some provinces/territories have been grouped together by Statistics Canada.
Episode week	Derived by Statistics Canada from EpisodeDate (not available on this dataset). Episode date is derived based on the earliest of the following dates: symptom onset date, specimen collection date, laboratory testing date, date reported to the province or territory, or date reported to PHAC. 0 represents the first days of the year leading up to, but not including the first Sunday. 1 represents the first full week of the year, beginning on the first Sunday, and so on.
Episode week group	Derived by Statistics Canada from Episode week. Indicates when multiple episode weeks have been grouped together to protect confidentiality. Refer to Appendix VI.
Episode year	Derived by Statistics Canada from EpisodeDate (not available on this dataset). Episode date is derived based on the earliest of the following dates: symptom onset date, specimen collection date, laboratory testing date, date reported to the province or territory, or date reported to PHAC.
Gender	Derived from the Gender variable received from PHAC (not available on this dataset). Where available, gender data was used; when gender data was unavailable, sex data was used. Missing values and ''Other'' were assigned to 'Not Stated'.
Age group	
Occupation	Derived by Statistics Canada from multiple variables (not available in this dataset). Healthcare workers include those with and without direct patient contact. Healthcare workers include any role in a private or public health care setting, including employee, volunteer, student. Long-term care residents may include residents of senior's homes, assisted living facilities, and retirement communities, as well as nursing homes. Long-term care facilities may be privately run or under provincial authority. Occupations with low frequencies have been categorized with the "Other".
Asymptomatic	Derived from the symptoms. If no symptoms were experienced, then asymptomatic is yes. If any symptoms were experienced then asymptomatic is no.
Onset week of symptoms	Derived from historic symptom data and asymptomatic variable. If any symptoms were reported then asymptomatic is no. 0 represents the first days of the year leading up to, but not including the first Sunday. 1 represents the first full week of the year, beginning on the first Sunday, and so on.
Onset year of symptoms	Derived by Statistics Canada from OnsetDate (not available on this dataset).
Hospital status	Routine updates on health outcome status are not made uniformly across Canada, and therefore the data may underestimate the number of hospitalizations, admissions to intensive care units, deaths and recoveries.
Resolved	
Resolution week	Derived by Statistics Canada from Resolution Date (not available on this dataset). 0 represents the first days of the year leading up to, but not including the first Sunday. 1 represents the first full week of the year, beginning on the first Sunday, and so on.
Resolution year	Derived by Statistics Canada from Resolution Date (not available on this dataset).
Death	Refer to the comment in “Hospital Status”.
Transmission	Domestic acquisition – Contact of COVID case: Includes cases who reported having close contact with a confirmed or probable COVID-19 case in the 14 days prior to symptom onset. Domestic acquisition– Contact with traveler: Includes cases who reported having close contact with a symptomatic person who had traveled to an affected area in the 14 days prior to their illness onset. Domestic acquisition – Unknown source: Includes cases who had not travelled, and 1) who had reported no contact with a COVID-19 case or symptomatic traveller, or 2) whose information on contact with a case or contact with a symptomatic traveler was unknown or missing. International travel: Includes cases who reported having travelled outside of their province / territory of residence or outside of Canada within the 14 days prior to symptom onset. Information pending: Includes cases for which information on contact with a case, contact with a symptomatic traveler, and travel history were all missing or unknown.

Appendix III - Data dictionary, Source, Format and Answer Categories

Concept	Source	Format	Answer Categories
Case identifier number	Statistics Canada	8.0	Continuous value from 1 to 99999999
Region	Public Health Agency of Canada	1.0	1=Atlantic (New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland and Labrador), 2=Quebec, 3=Ontario and Nunavut, 4=Prairies (Manitoba, Saskatchewan, Alberta) and the Northwest Territories, 5=British Columbia and Yukon
Episode week	Public Health Agency of Canada	2.0	Continuous value from 0 to 52, 99=Not stated
Episode week group	Public Health Agency of Canada	2.0	Refer to Appendix VI.
Episode year	Public Health Agency of Canada	2.0	20=2020, 21=2021, 99=Not stated
Gender	Public Health Agency of Canada	1.0	1=Male, 2=Female, 9=Not stated/Other
Age group	Public Health Agency of Canada	2.0	1=0-19, 2=20-29, 3=30-39, 4=40-49, 5=50-59, 6=60-69, 7=70-79, 8=80+, 99=Not stated
Occupation	Public Health Agency of Canada	2.0	1=Health Care Worker, 2=School or daycare worker/attendee, 3=Long term care resident, 4=Other, 9=Not stated
Asymptomatic	Public Health Agency of Canada	1.0	1=Yes, 2=No, 9=Not Stated
Onset week of symptoms	Public Health Agency of Canada	2.0	Continuous value from 0 to 52, 99=Not stated or Not applicable
Onset year of symptoms	Public Health Agency of Canada	2.0	20=2020, 21=2021, 99=Not stated
Hospital status	Public Health Agency of Canada	1.0	1=Hospitalized - ICU, 2=Hospitalized - Non-ICU, 3=Not Hospitalized, 9=Not stated/Unknown
Resolved	Public Health Agency of Canada	1.0	1=Yes, 2=No, 9=Not Stated/Unknown
Resolution week	Public Health Agency of Canada	2.0	Continuous value from 0 to 52, 99=Not stated/Not applicable
Resolution year	Public Health Agency of Canada	2.0	20=2020, 21=2021, 99=Not stated
Death	Public Health Agency of Canada	1.0	1=Yes, 2=No, 9=Not Stated
Transmission	Public Health Agency of Canada	1.0	1=Domestic Acquisition: “Contact of COVID Case” or “Contact with traveler” or “Unknown Source”, 2=International Travel, 9=Not stated/Pending

Appendix IV – Information on the Week Variables for the year 2020

Week	Description	Week	Description	Week	Description
0	Week of December 29th	18	Week of May 3rd	36	Week of September 6th
1	Week of January 5th	19	Week of May 10th	37	Week of September 13th
2	Week of January 12th	20	Week of May 17th	38	Week of September 20th
3	Week of January 19th	21	Week of May 24th	39	Week of September 27th
4	Week of January 26th	22	Week of May 31st	40	Week of October 4th
5	Week of February 2nd	23	Week of June 7th	41	Week of October 11th
6	Week of February 9th	24	Week of June 14th	42	Week of October 18th
7	Week of February 16th	25	Week of June 21th	43	Week of October 25th
8	Week of February 23rd	26	Week of June 28th	44	Week of November 1st
9	Week of March 1st	27	Week of July 5th	45	Week of November 8th
10	Week of March 8th	28	Week of July 12th	46	Week of November 15th
11	Week of March 15th	29	Week of July 19th	47	Week of November 22nd
12	Week of March 22nd	30	Week of July 26th	48	Week of November 29th
13	Week of March 29th	31	Week of August 2nd	49	Week of December 6th
14	Week of April 5th	32	Week of August 9th	50	Week of December 13th
15	Week of April 12th	33	Week of August 16th	51	Week of December 20th
16	Week of April 19th	34	Week of August 23rd	52	Week of December 27th
17	Week of April 26th	35	Week of August 30th		

Appendix V – Information on the Week Variables for the year 2021

Week	Description	Week	Description	Week	Description
0	NA	18	Week of May 2nd	36	Week of September 5th
1	Week of January 3rd	19	Week of May 9th	37	Week of September 12th
2	Week of January 10th	20	Week of May 16th	38	Week of September 19th
3	Week of January 17th	21	Week of May 23rd	39	Week of September 26th
4	Week of January 24th	22	Week of May 30th	40	Week of October 3rd
5	Week of January 31st	23	Week of June 6th	41	Week of October 10th
6	Week of February 7th	24	Week of June 13th	42	Week of October 17th
7	Week of February 14th	25	Week of June 20th	43	Week of October 24th
8	Week of February 21st	26	Week of June 27th	44	Week of October 31st
9	Week of February 28th	27	Week of July 4th	45	Week of November 7th
10	Week of March 7th	28	Week of July 11th	46	Week of November 14th
11	Week of March 14th	29	Week of July 18th	47	Week of November 21st
12	Week of March 21st	30	Week of July 25th	48	Week of November 28th
13	Week of March 28th	31	Week of August 1st	49	Week of December 5th
14	Week of April 4th	32	Week of August 8th	50	Week of December 12th
15	Week of April 11th	33	Week of August 15th	51	Week of December 19th
16	Week of April 18th	34	Week of August 22nd	52	Week of December 26th
17	Week of April 25th	35	Week of August 29th		

Appendix VI - Information on the Episode Week Group Indicator

Episode week group	Region(s)	Episode weeks grouped	Grouped with episode week	Episode year
0	All regions	No grouping	NA	NA
1	All regions	0 to 8	8	20
2	Atlantic	23, 24 and 25	23	20
3	Atlantic	26, 27, 28, 29, 30 and 31	28	20
4	Atlantic	32, 33, 34 and 35	32	20
5	Atlantic	36, 37, 38 and 39	39	20
6	Atlantic	10, 11 nad 12	11	21
7	Atlantic	26, 27, 28, 29 and 30	26	21