

# User Guide and Data Dictionary for Preliminary COVID-19 Data

Release date: November 12, 2022  
Updated: January 12, 2024



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](http://www.statcan.gc.ca).

You can also contact us by

**Email at** [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.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](http://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.

Published by authority of the Minister responsible for Statistics Canada

© His Majesty the King in Right of Canada, as represented by the Minister of Industry, 2024

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An [HTML version](#) is also available.

*Cette publication est aussi disponible en français.*

---

## Table of contents

<b>1. Background.....</b>	<b>4</b>
<b>2. Objectives of the Preliminary dataset on confirmed cases of COVID-19 .....</b>	<b>4</b>
<b>3. Coverage of the Preliminary dataset on confirmed cases of COVID-19.....</b>	<b>4</b>
<b>4. Content of the Preliminary dataset on confirmed cases of COVID-19<sup>1</sup> .....</b>	<b>5</b>
<b>5. Limitations.....</b>	<b>6</b>
<b>6. Data quality concerns .....</b>	<b>7</b>
<b>7. References .....</b>	<b>7</b>
<b>Appendix I - Data dictionary, Concept, Variable Name, Section on the Form, Description and Universe .....</b>	<b>8</b>
<b>Appendix II - Data dictionary, Notes and Limitations .....</b>	<b>9</b>
<b>Appendix III - Data dictionary, Source, Format and Answer Categories .....</b>	<b>10</b>
<b>Appendix IV – Information on the Week Variables for the Year 2020.....</b>	<b>11</b>
<b>Appendix V – Information on the Week Variables for the Year 2021.....</b>	<b>12</b>
<b>Appendix VI – Information on the Week Variables for the Year 2022.....</b>	<b>13</b>
<b>Appendix VII – Information on the Week Variables for the Year 2023.....</b>	<b>14</b>
<b>Appendix VIII – Information on the Week Variables for the Year 2024.....</b>	<b>15</b>
<b>Appendix IX - Information on the Episode Week Group Indicator.....</b>	<b>16</b>

# 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 had not been identified before the first cases were reported in Wuhan, China, in December 2019. The first confirmed cases in Canada appeared in January 2020.

In Canada, the 10 provinces and 3 territories are providing the Public Health Agency of Canada (PHAC) information on COVID-19 cases on a routine 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 from 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 provincial or territorial public health authorities to PHAC. The governments of Canada and the provinces and territories agreed on a common Case Report Form (CRF)<sup>1</sup> 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 routinely by each jurisdiction. Discrepancies are due to factors such as delays in reporting, or variability in reporting cut-offs. Due to these discrepancies, the 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 and deaths.

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.

Variables related to the resolution status of cases were removed from the data file in June 2022. Many provinces and territories were unable to determine the resolution status of a case in alignment with the national definition of a resolved COVID-19 infection.<sup>2</sup>

The variables transmission, asymptomatic, occupation, onset year of symptoms and onset month of symptoms were removed from the data file in January 2023. Most provinces or territories were no longer reporting these variables meaning they are incomplete and no longer representative, creating challenges for interpretation as a result the quality of these variables are no longer deemed sufficient for release.

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<sup>1</sup>

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 derived variables have been created:
  - ▶ Dates (Earliest Date)
    - Earliest Date is used to derive Episode week, Episode week group and Episode year.
    - All dates were converted to weeks, i.e., Episode week, Episode week group.
    - All cases with an earliest date prior to February 23<sup>rd</sup>, 2020 (the first day of the 8<sup>th</sup> week of 2020) were grouped with the cases in week 8.
    - 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 IX will indicate which weeks have been grouped together.

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,<sup>2</sup> that is:

- The detection of at least 1 specific gene target by a validated laboratory-based nucleic acid amplification test (NAAT) assay performed at a community, hospital, or reference laboratory, or
- the detection of at least 1 specific gene target by a validated point-of-care NAAT that has been deemed acceptable to provide a final result, or
- seroconversion or diagnostic rise (at least 4-fold or greater from baseline) in viral specific antibody titre in serum or plasma using a validated laboratory-based serological assay for SARS-CoV-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.

Due to changes in COVID-19 testing policies in many jurisdictions, which were in response to a surge in demand for laboratory testing starting in December 2021, cases may be skewed towards populations deemed high priority for laboratory testing, and case counts will underestimate the total burden of COVID-19 in the population.

Furthermore, following the surge in demand for laboratory testing, COVID-19 rapid antigen tests (RATs) became increasingly available for public use. While positive laboratory tests have a reporting mechanism to provincial public health authorities, RATs do not, and are therefore not captured within this data set. The availability of RATs for public use limited demand for laboratory testing, resulting in case counts further underestimating of the total burden of COVID-19 in the population.

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

- 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.
- Starting in December 2021, cases may be skewed towards populations deemed high priority for laboratory testing, and case counts will underestimate the total burden of COVID-19 in the population.

## 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 and deaths.

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

## 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) <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) <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 IX.	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
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
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



**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 Earliest Date (not available on this dataset). Earliest 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 IX.
Episode year	Derived by Statistics Canada from Earliest Date (not available on this dataset). Earliest 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	
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.
Death	Derived by Statistics Canada from COVIDDeath (not available on this dataset). 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 and deaths.

**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 IX.
Episode year	Public Health Agency of Canada	2.0	20=2020, 21=2021, 22=2022, 23=2023, 24=2024, 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
Hospital status	Public Health Agency of Canada	1.0	1=Hospitalized - ICU, 2=Hospitalized - Non-ICU, 3=Not Hospitalized, 9=Not stated/Unknown
Death	Public Health Agency of Canada	1.0	1=Yes, 2=No, 9=Not Stated

**Appendix IV – Information on the Week Variables for the Year 2020**

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

**Appendix V – Information on the Week Variables for the Year 2021**

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

**Appendix VI – Information on the Week Variables for the Year 2022**

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

**Appendix VII – Information on the Week Variables for the Year 2023**

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

**Appendix VIII – Information on the Week Variables for the Year 2024**

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

**Appendix IX - 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 and 12	11	21
7	Atlantic	26, 27, 28, 29 and 30	26	21