

COVID-19 Case Surveillance Public Use Data

Multiple/Other, Non-Hispanic
view data

Visualize

Export

API

...

Case Surveillance

The COVID-19 case surveillance system database includes patient-level data reported to U.S. states and autonomous reporting entities, including New York City and the District of Columbia (D.C.), as well as U.S. territories and states. On April 5, 2020, COVID-19 was added to the Nationally Notifiable Condition List and classified as “immediately notifiable, urgent (within 24 hours)” by a Council of State and Territorial Epidemiologists (CSTE) Interim Position Statement (Interim-20-ID-01). The statement also recommended that all states and territories enact laws to make COVID-19 reportable in their jurisdiction, and that jurisdictions conducting surveillance should submit case notifications to CDC. COVID-19 case surveillance data are collected and reported voluntarily to CDC’s COVID-19 Response.

Updated
September 30, 2020
Data Provided by
CDC Case Surveillance Task Force

These deidentified data include demographic characteristics, exposure history, disease severity indicators and outcomes, clinical data, laboratory diagnostic test results, and comorbidities. All data elements can be found on the COVID-19 case report form located at www.cdc.gov/coronavirus/2019-ncov/downloads/pui-form.pdf.

The Case Surveillance Task Force and Surveillance Review and Response Group (SRRG) within CDC’s COVID-19 Response provide stewardship for datasets that support the public health community’s access to COVID-19 data while protecting patient privacy.

COVID-19 case reports have been routinely submitted using standardized case reporting forms. On April 5, 2020, CSTE released an Interim Position Statement with national surveillance case definitions for COVID-19 included. Current versions of these case definitions are available here: <https://wwwn.cdc.gov/nndss/conditions/coronavirus-disease-2019-covid-19/>. All cases reported on or after were requested to be reported by public health departments to CDC using the standardized case definitions for lab-confirmed or probable cases. On May 5, 2020, the standardized case reporting form was revised. Implementation of case reporting using this new form is ongoing among U.S. states and territories.

The COVID-19 case surveillance data are dynamic; case reports can be modified at any time by the reporting jurisdiction as new information becomes available (i.e., data are subject to change). Furthermore, reporting jurisdictions may report cases late. Version updates to the detailed and limited datasets will be available for request once a month. The datasets will include all cases with an initial report date of case to CDC at least 14 days prior to the creation of the previously updated datasets. This 14 day lag will allow case reporting to be stabilized and ensure that time-dependent outcome data, including death, are accurately captured.

CDC’s Case Surveillance Task Force routinely performs data quality assurance procedures (i.e., ongoing corrections and logic checks to address data errors). To date, the following data cleaning steps have been implemented:

Questions that have been left unanswered (blank) on the case report form are re-classified to a Missing value, if applicable to the question. For example, in the question “Was the patient hospitalized?”, where the possible answer choices include “Yes”, “No”, or “Unknown”, the missing value is re-coded to Missing if the respondent did not answer the question.

Logic checks are performed for date data. If an illogical date has been provided, CDC reviews the data with the reporting jurisdiction. For example, if a symptom onset date that is in the future is reported to CDC, this value is set to null until the reporting jurisdiction updates this information appropriately.

The initial report date of the case to CDC is intended to be completed by the reporting jurisdiction when data are submitted. If blank, this variable is completed using the date the data file was first submitted to CDC.

Additional data quality processing to recode free text data are ongoing. Data on symptoms, race and ethnicity, and healthcare worker status have been prioritized.

To prevent release of data that could be used to identify persons, data cells are suppressed for low frequency (<5) records and indirect identifiers (date of first positive specimen). Suppression includes uncommon combinations of demographic characteristics (sex, age group, race/ethnicity). Suppressed values are re-coded to the *NA* answer option.

Restricted use detailed data set is available to researchers willing to enter a "Registration Information and Data Use Restrictions Agreement (RIDURA)" by emailing a request to eoevent394@cdc.gov.

COVID-19 data will be made available to the public as summary or aggregate count files, including total counts of cases and deaths by state and by county. These and other data on COVID-19 are available from multiple public locations:

<https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html>

<https://www.cdc.gov/covid-data-tracker/index.html>

<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>

<https://www.cdc.gov/coronavirus/2019-ncov/php/open-america/surveillance-data-analytics.html>

Less

Featured Content Using this Data

Demographic Trends of COVID-19 cases and death...

External Content

CDC is working with states to provide more information on race/ethnicity for reported case. The percent of reported cases that include ra...

About this Dataset

Mute Dataset

Updated
September 30, 2020

Data Last Updated
September 30, 2020

Metadata Last Updated
September 30, 2020

Date Created
May 15, 2020

Views
339K

Downloads
18K

Data Provided by
CDC Case Surveillance Task Force

Dataset Owner
Data.CDC.gov Administrator

Contact Dataset Owner

Common Core

Publisher	CDC
Contact Name	Surveillance Review and Response Group
Contact Email	eoevent394@cdc.gov
Suggested Citation	Centers for Disease Control and Prevention, COVID-19 Response. COVID-19 Case Surveillance Public Data Access, Summary, and Limitations (version date: September 30, 2020).
Bureau Code	009:20
Program Code	009:020

Data Quality

Geographic Coverage	US
Temporal Applicability	2020-01-01/2020-09-15
Update Frequency	Monthly

Topics

Category	Case Surveillance
Tags	covid-19 , cases , surveillance , covid , covid19

Licensing and Attribution

License	Public Domain U.S. Government
Source Link	https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/index.html

Show Less

What's in this Dataset?

Rows
4.48M

Columns
11

Each row is a
Deidentified Patient

Columns in this Dataset

Column Name	Description	Type
cdc_report_dt	Initial case report date to CDC	Date & Time
pos_spec_dt	Date of first positive specimen collection	Date & Time
onset_dt	Symptom onset date, if symptomatic	Date & Time
current_status	Case Status: Laboratory-confirmed case; Probable case	Plain Text
sex	Sex: Male; Female; Unknown; Other	Plain Text
age_group	Age Group: 0 - 9 Years; 10 - 19 Years; 20 - 39 Years; 40 - 49 ...	Plain Text
Race and ethnicity (combined)	Race and ethnicity (combined): Hispanic/Latino; American I...	Plain Text
hosp_yn	Hospitalization status	Plain Text
icu_yn	ICU admission status	Plain Text
death_yn	Death status	Plain Text
medcond_yn	Presence of underlying comorbidity or disease	Plain Text

Show Less

Table Preview

View Data

Create Visualization

cdc_re...	pos_sp...	onset_dt	current...	sex	age_gr...	Race a...	hosp_yn	icu_yn	death_...	medco...
2020/03/22	2020/03/15	2020/03/11	Laboratory...	Female	0 - 9 Years	Multiple/Ot...	Yes	Yes	No	Yes
2020/03/23	2020/03/15		Laboratory...	Female	0 - 9 Years	Unknown	Unknown	Unknown	Unknown	Unknown
2020/03/15	2020/03/15		Laboratory...	Female	0 - 9 Years	Unknown	Missing	Missing	Missing	Missing
2020/03/15	2020/03/15		Laboratory...	Female	0 - 9 Years	Unknown	Missing	Missing	Missing	Missing
2020/03/15	2020/03/15		Laboratory...	Female	0 - 9 Years	Unknown	Missing	Missing	Missing	Missing
2020/03/03	2020/03/15	2020/03/03	Laboratory...	Female	0 - 9 Years	Unknown	No	Unknown	No	No
2020/03/20	2020/03/15		Laboratory...	Female	0 - 9 Years	Unknown	Missing	Missing	Missing	Unknown
2020/03/20	2020/03/15		Laboratory...	Female	0 - 9 Years	Unknown	Missing	Missing	Missing	Unknown
2020/03/24	2020/03/15		Laboratory...	Male	0 - 9 Years	Hispanic/La...	Missing	Missing	Missing	Unknown
2020/03/15	2020/03/15	2020/03/30	Laboratory...	Male	0 - 9 Years	Hispanic/La...	No	No	No	No
2020/03/13	2020/03/15	2020/03/13	Laboratory...	Male	0 - 9 Years	Hispanic/La...	No	Missing	No	Missing
2020/03/14	2020/03/15	2020/03/14	Laboratory...	Female	0 - 9 Years	Black, Non-...	No	Unknown	Unknown	No
2020/03/18	2020/03/15	2020/03/13	Laboratory...	Male	0 - 9 Years	Multiple/Ot...	No	No	No	No
2020/03/14	2020/03/15	2020/03/14	Laboratory...	Female	0 - 9 Years	Hispanic/La...	No	Missing	No	Missing

< Previous Next >

Showing Deidentified Patients 1 to 14 out of 4,481,062

Related Content Using this Data

COVID-19 Case Surveillance
Public Use Data Profile

September 30, 2020 9,310 Views

Summary of variables.

The COVID-19 case surveillance system dat...