



## COVID-19

# FAQ: COVID-19 Data and Surveillance

## Frequently Asked Questions

Updated Apr. 14, 2021 [Print](#)

## National COVID-19 Case Surveillance

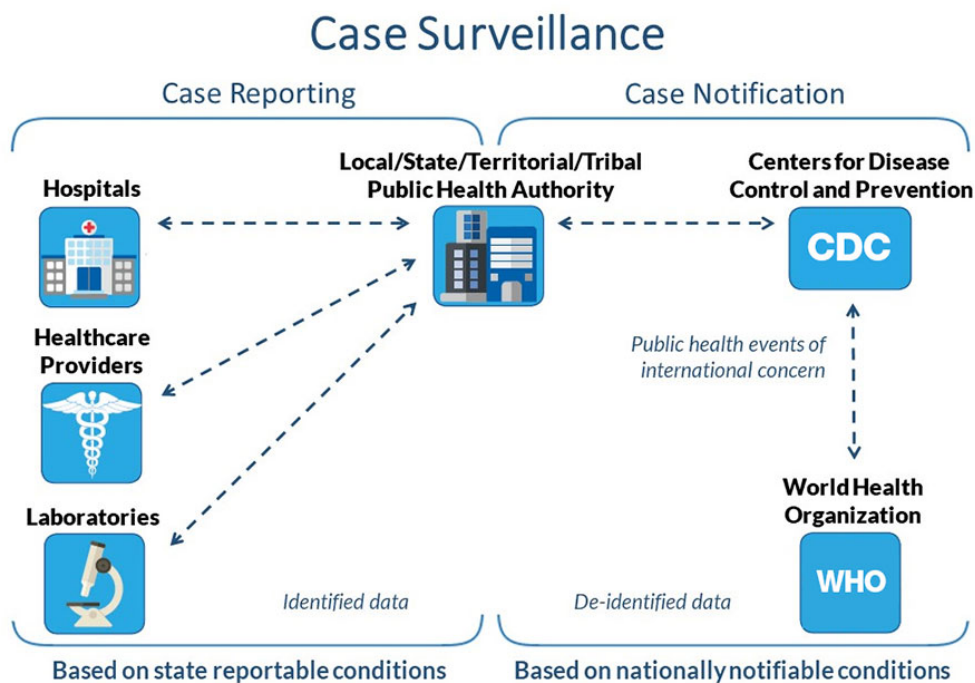
### What is national case surveillance?

To protect Americans from serious infectious diseases and other health threats, public health authorities conduct national case surveillance to monitor more than 120 diseases and conditions. For these conditions, public health departments collect information on individuals with the infection in a population, which is known as case surveillance. One goal of case surveillance is to provide information needed for taking public health action to prevent cases and spread of disease; another is to control outbreaks. Case surveillance is especially important for new diseases, such as COVID-19, to understand the similarities and differences among cases, including:

- Demographic, clinical, and epidemiologic characteristics;
- Exposure and contact history; and
- Course of clinical illness and care received.

During the COVID-19 response, state and jurisdictional health departments voluntarily send case data to CDC using the [National Notifiable Diseases Surveillance System](#). To protect individuals' privacy, COVID-19 case data are sent to CDC without personal identifiers, such as names or home addresses. A national standardized [case definition](#) is used to define confirmed, probable, and suspect cases and deaths.

Unlike data collected for clinical trials and research studies, in which scientists comprehensively measure and follow the health status of patients, national case surveillance data focus on capturing demographic and risk factor information about people with COVID-19.

How is COVID-19 case information collected and reported? 

 [View Larger](#)

The process for reporting, collecting, and analyzing disease data is called a *data supply chain*. Under state disease reporting laws, hospitals, healthcare providers, and laboratories must report confirmed or probable COVID-19 cases and deaths to state or local health departments. These laws are designed to help health departments quickly identify outbreaks and control the spread of disease. The figure below illustrates how data are transferred for case reporting (from hospitals, healthcare providers, and laboratories to local, state, regional, or territorial public health) and how data move for case notification (from state or territorial public health departments to CDC). These two steps of information flow make up national case surveillance. While case reporting is mandatory under state reportable disease laws, case *notification* from state and local health departments to CDC is voluntary and includes deidentified data.

Using the [National Notifiable Diseases Surveillance System](#), health departments voluntarily send COVID-19 case and death data to CDC that do not include personally identifiable information. Because COVID-19 has been designated as a public health emergency of international concern, CDC reports national case surveillance data to the World Health Organization under [International Health Regulations](#) (2005). CDC also publishes deidentified COVID-19 national case surveillance data at [data.cdc.gov](https://data.cdc.gov), with additional privacy protections in place for public use.

## What data sources does CDC use for COVID-19 case surveillance?

To obtain timely and detailed data on COVID-19 cases in the United States, CDC uses two data sources. The first data source is an **aggregate count** based on a robust, multistep process to collect data and confirm the case and death numbers with jurisdictions daily:

- A CDC data team collects information from jurisdictions' websites, and a separate CDC data team double-checks the information collected.
- CDC then shares the data back with the jurisdictions for confirmation or corrections.
- CDC reconciles any differences and posts the finalized information to a CDC website.

This process is collaborative, with CDC and the jurisdictions working together to ensure the accuracy of the COVID-19 case and death numbers published on CDC's website. Aggregate counts provide the most up-to-date validated numbers on cases and deaths; CDC may retrospectively update the counts after posting based on any updated information from jurisdictions.

The second data source involves **line-level data for each case**, which provide additional information about whether the patient died and other details such as age and race and ethnicity. CDC receives the line-level data primarily from state health departments without personal identifiers such as names or home addresses. Because it can be time-consuming for jurisdictions to collect the additional information, these data can lag behind the aggregate counts. Although CDC receives this information for most cases, it does not receive it for all cases.

## How complete are the data that CDC receives for COVID-19 cases?

The COVID-19 pandemic has put unprecedented demands on the public health data supply chain. In many states, the large number of COVID-19 cases has severely strained the ability of hospitals, healthcare providers, and laboratories to report cases with complete demographic information, such as race and ethnicity. The unprecedented volume of cases has also limited the ability of state and local health departments to conduct thorough case investigations and collect all requested case data.

As a result, many COVID-19 case notifications submitted to CDC do not have complete information on patient demographics; signs and symptoms of illness; underlying health conditions; characteristics of hospitalizations such as ventilator use; clinical outcomes; exposures; and factors that may put people at higher risk for severe disease. Because it can be time-consuming for jurisdictions to collect the additional information, these data can lag behind the aggregate counts. Because of missing data, analyses of these data elements are likely an underestimate of the true occurrence.

Most states have demographic factors like age and sex for the majority of reported cases. With thousands of cases being reported, however, completeness of these elements is unlikely to improve in the immediate future for some jurisdictions.

Because the racial and ethnic composition of the U.S. population varies by geographic area, comparisons of COVID-19 case information should consider the population of each geographic area. Additionally, because completeness of race and ethnicity information may vary by state or geographic area and other patient factors, such as severity of illness, CDC's case data may not be generalizable to the entire U.S. population.

## How does CDC use COVID-19 national case surveillance data?

Case surveillance provides information on the characteristics of a disease within a population, usually through laboratory confirmation of cases using a standard [case definition](#). CDC uses national case surveillance to:

- Track the spread of COVID-19 around the country to identify areas of concern and inform state decision makers;
- Help state and local public health departments better control COVID-19 by evaluating trends in case demographics, exposures, and outcomes to identify those groups most at risk, such as healthcare workers, racial and ethnic minorities, older adults, and people with certain underlying health conditions; and
- Analyze exposure information and health outcomes among COVID-19 patients to develop guidance for the public, at-risk groups, and healthcare providers.

## What are some of the challenges with national case surveillance for COVID-19?

The COVID-19 pandemic has put unprecedented strain on the public health data supply chain. In many states, the large number of COVID-19 cases has severely strained the ability of hospitals, healthcare providers, and laboratories to report cases with complete demographic information, such as race and ethnicity. The unprecedented volume of cases has also limited the ability of state and local health departments to conduct thorough case investigations and collect all requested case data. As a result, many COVID-19 case notifications submitted to CDC do not have complete information on patient demographics, clinical outcomes, exposures, and factors that may put people at higher risk for severe disease.

National case surveillance data are constantly changing. For instance, as new information is gathered about previously reported cases, health departments provide updated data to CDC. As more information and data become available, analyses might find changes in surveillance data and trends during a previously reported time window.

A key challenge with case reporting is that people who are infected with the virus that causes COVID-19 may have mild or no symptoms. These people might not have sought testing or health care and are, therefore, less likely to be reported as cases. Similarly, cases in people who have had severe outcomes, such as hospitalization, intensive care unit (ICU) admission, and death, are more likely to be reported than cases in people with less severe illnesses. These challenges result in limitations when analyzing and interpreting the data.

## Does CDC update COVID-19 data on cases and deaths if jurisdictions submit or post the data late?

Yes. The following information details how CDC handles jurisdictions' historical data.

- If a jurisdiction provides historical data that include dates for the related events (for example, cases and deaths), CDC incorporates the historical data into the jurisdiction's cumulative data as soon as possible.
- If a jurisdiction provides historical data with no information about applicable dates for the event (for example, cases and deaths), CDC includes the historical data in the cumulative counts, but omits the historical data from other metrics until the jurisdiction provides the dates associated with the events (cases, deaths), at which point they will be applied to the appropriate dates.
- When a jurisdiction responds to CDC's request for dates to assign to their historical data:
  - If they are unable to provide specific dates but can provide a date range for the historical data, these data are equally distributed across the date range provided by the jurisdiction.
  - If a jurisdiction does not know what date or date range applies to its historical data, CDC equally distributes the data across the first date the jurisdiction began submitting data to CDC through the date the data were first received.

## What is CDC doing to improve national case surveillance and reporting?

CDC continues to work with state, local, and territorial health departments to accelerate reporting of national case surveillance data, improve data quality, and gather complete information about all COVID-19 cases.

CDC is working with healthcare providers, electronic health record developers, laboratories, and state and local health departments to [modernize](#) disease surveillance by automating the generation and transmission of case reports from the electronic health record to public health agencies for review and action for the COVID-19 response.

For example, expanded use of electronic case reporting, which make the submission of information from healthcare providers to public health departments seamless and automated, will reduce the burden of manually reporting COVID-19 cases, increase timeliness of reporting, and improve data completeness by pulling data directly from the medical record.

### Additional resources

- [CDC COVID Data Tracker](#)
- [CDC's National Notifiable Diseases Surveillance System](#)
- [Electronic Case Reporting for COVID-19](#)
- [CSTE Case Definition for COVID-19](#)
- [International Health Regulations \(2005\)](#)
- [Public Health Surveillance in the United States: Evolution and Challenges, July 2012](#)
- [Modernizing Centers for Disease Control and Prevention Informatics Using Surveillance Data Platform Shared Services !\[\]\(a9a7cf821bf949be41db724492f295be\_img.jpg\) , March–April 2018](#)
- [CDC's Vision for Public Health Surveillance in the 21st Century, July 2012](#)
- Centers for Disease Control and Prevention (CDC). [Introduction to Public Health](#). In: [Public Health 101 Series](#). Atlanta, GA: U.S. Department of Health and Human Services, CDC; 2014.
- [CDC MMWR Novel Coronavirus Reports](#)

---

# CDC COVID-19 Surveillance



## How does CDC collect COVID-19 surveillance data?

**Public health surveillance** is the ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of public health practice.

For surveillance of COVID-19 and the virus that causes it, SARS-CoV-2, CDC is using multiple surveillance systems in collaboration with state, local, territorial, academic, and commercial partners to monitor COVID-19 in the United States. COVID-19 surveillance draws from a combination of data sources using existing influenza and viral respiratory disease surveillance, **syndromic surveillance**, case reporting, lab reporting, health care systems reporting, ongoing research platforms, and new surveillance systems designed to answer specific questions. Combined, the data from these systems create an updated picture of COVID-19's spread and its effects in the United States and are used to inform the U.S. national public health response to COVID-19.

## What are the goals for COVID-19 national surveillance?

- To monitor spread of COVID-19 in the United States
- To understand disease severity and the spectrum of illness due to COVID-19
- To understand risk factors for severe disease and transmission of COVID-19
- To monitor for changes in the virus that causes COVID-19
- To estimate disease burden due to COVID-19
- To produce data for forecasting COVID-19's spread and impact
- To understand how COVID-19 impacts the capacity of the U.S. healthcare system (for example, availability and shortages of key resources)

## How are COVID-19 data used?

COVID-19 data can be used to help public health professionals, policymakers, and healthcare providers monitor the spread of COVID-19 in the United States and support a better understanding of the spectrum of illness, the effectiveness of community intervention, and social disruptions associated with COVID-19 in the United States. These data help inform U.S. national, state, local, tribal, and territorial public health responses to COVID-19.

Detailed and accurate data will allow us to better understand and track the size and scope of the outbreak and strengthen prevention and response efforts.

## Where can I find information on COVID-19 cases, testing, hospitalizations, deaths, and other surveillance in the United States?

CDC provides this information on the [Cases, Data, & Surveillance](#) webpage. The following types of information are available on the webpage:

- **CDC COVID Data Tracker**
  - **CDC COVID Data Tracker** is CDC's home for COVID-19 data. It provides surveillance data from across the response, including hospitalizations, vaccinations, demographic information, and daily and cumulative case and death counts reported to CDC since January 21, 2020. Data found on Data Tracker are updated daily.
  - Topics covered on COVID Data Tracker include:
    - COVID-19 in Your Community
    - Vaccinations
    - Cases & Deaths
    - Demographic Trends
    - Health Care Settings
    - Genomic Surveillance
    - Testing and Seroprevalence
    - People at Increased Risk
- **COVID Data Tracker Weekly Review**
  - Key updates for the week (trends in cases, deaths, variants, laboratory testing, hospitalizations, and vaccinations)
  - Interpretive summaries for trends in key COVID-19 data
- **Hospitalization Surveillance Network (COVID-NET)**


- National hospitalization rates for COVID-19
  - Characteristics of people hospitalized with COVID-19 in the U.S.
- **COVID-19 Serology Surveillance**
  - Information on large-scale geographic, community-level, and special populations seroprevalence surveys (Results from these surveys will be posted as they become available.)
- **COVID-19 Data from the National Center for Health Statistics (NCHS)**
  - National Vital Statistic System's provisional death counts
  - Data on mental health and access to healthcare from the NCHS partnership with the U.S. Census Bureau on the Household Pulse Survey (includes indicators of anxiety and depression based on reported frequency of symptoms during the last 7 days)
- **Patient Impact and Hospital Capacity Pathway**
  - Estimated percentage of inpatient beds occupied by all patients, by state
  - Estimated percentage of inpatient beds occupied by COVID-19 patients, by state
  - Estimated percentage of ICU beds occupied by all patients, by state

---

## Understanding the Data

### What is a COVID-19 case?

A COVID-19 case includes [confirmed and probable cases and deaths](#). The case classifications for COVID-19 are described in an updated interim [COVID-19 position statement](#) and case definition issued by the Council of State and Territorial Epidemiologists on August 5, 2020. Although this updated case definition includes three case classifications (suspect, probable, and confirmed), CDC case counts exclude [suspect cases and deaths](#).

A previous [COVID-19 position statement](#)  issued by CSTE on April 5, 2020, included a case definition and made COVID-19 a [nationally notifiable disease](#). A notifiable disease or condition is one for which regular, frequent, and timely information regarding individual cases is considered necessary to prevent and control the disease or condition.

## What is a COVID-19 probable case?

A [probable case or death](#) is defined as any one of the following:

- Meets [clinical criteria](#) AND [epidemiologic linkage](#) with no confirmatory laboratory testing performed for SARS-CoV-2;
- Meets [presumptive laboratory evidence](#);
- Meets [vital records criteria](#) with no confirmatory laboratory evidence for SARS-CoV-2.

## How is CDC handling probable cases and deaths from jurisdictions?

CDC applies a standard protocol for data reporting across jurisdictions in line with the [2020 Interim Case Definition](#) outlined in CSTE's Position Statement (approved August 5, 2020), such that CDC includes jurisdictions' reported confirmed and probable cases as "cases," and includes jurisdictions' reported confirmed and probable deaths as "deaths." For example, jurisdictions' reports of new confirmed and new probable cases are summed to reflect new cases; the process is similar for deaths. Jurisdictions' probable cases and deaths are also included in their cumulative counts.

## Why do the number of new cases change over time and by geographical region?

The virus that causes COVID-19 [spreads](#) very easily and sustainably between people. The more closely people interact with others and the longer that interaction, the higher the risk of COVID-19 spread. Practicing preventive actions such as avoiding close contact, wearing face coverings, washing hands often, and cleaning and disinfecting prevent the spread of COVID-19. Differences in community characteristics and changes in preventive behavior can result in increases or decreases of cases over time and geographical area.

## Why are the death counts on the Cases and Deaths by State tab on the COVID-19 Data Tracker different from the provisional death counts on the COVID-19 webpage?

The COVID-19 death count shown on the [Cases and Deaths by State tab](#) on the COVID-19 Data Tracker includes deaths reported daily by state, local, and territorial health departments. This count reflects the most up-to-date information received by CDC based on preliminary reporting from health departments.

In contrast, [provisional COVID-19 death counts](#) from the National Center for Health Statistics (NCHS) are updated Monday through Friday with information collected from death certificates. These data represent the most accurate death counts. However, because it can take several weeks for death certificates to be submitted and processed, there is on average a delay of 1–2 weeks before they are reported. Therefore, the provisional death counts may not include all deaths that occurred during a given time period, especially for more recent time periods. Death counts from earlier weeks are continually revised and may increase or decrease as new and updated death certificate data are received. Provisional COVID-19 death counts may therefore differ from those on other published sources, such as media reports or the [COVID-19 Data Tracker webpage](#).

## What does the mortality rate or death rate mean?

The mortality rate is the number of people who died due to COVID-19 divided by the total number of people in the population. Since this is an ongoing outbreak, the mortality rate can change daily.

## Why do some case numbers reported by state health departments, Johns Hopkins, and the World Health Organization (WHO) sometimes differ from what is posted on CDC's website?

Case numbers reported on other websites may differ from what is posted on CDC's website because CDC's overall case numbers are validated through a confirmation process with each jurisdiction. The process used for finding and confirming cases displayed by other reporting jurisdictions may differ. Differences between reporting jurisdictions and CDC's website may occur due to the timing of reporting and website updates.

## What are the limitations in using case surveillance data to understand the epidemiology of COVID-19?

Case surveillance data are useful for tracking national trends in disease incidence (the number of new cases of a disease in a population at a certain time period). Limitations of using case surveillance data to understand the epidemiology (who, what, where, when, how) of COVID-19 include the following:

First, case surveillance data do not represent the true burden of COVID-19 in the United States. Many people infected with the virus that causes COVID-19 do not seek medical care or get tested. The information collected might be limited if people are unavailable or unwilling to provide additional information or if medical records are unavailable for data extraction.

Second, most of the case reports captured by health departments are based on laboratory reports that usually contain limited information on the patient. Because of the volume of cases, most health departments are unable to conduct investigations of every case to obtain additional information. Because of this, most case reports are missing data on patient demographics, symptoms, underlying health conditions, characteristics of hospitalizations such as ventilator use, and other factors such as recent travel history. Because of missing data, analyses of these data elements are likely an underestimate of the true occurrence.

Third, it is difficult to capture asymptomatic cases through case surveillance. People who are asymptomatic are unlikely to seek testing unless they are identified through active screening (e.g., contact tracing), and investigation of symptomatic people is prioritized.

When disease volume is high and a limited number of data elements are captured on each reported case, case surveillance data can be used to assess population burden, track the spread of the disease, monitor increases and decreases in cases in association with mitigation strategies, and study selected demographics such as age, sex, race and ethnicity, and geography. Clinical details and other characteristics about people with COVID-19 can be better assessed through special studies. CDC conducts these [special epidemiologic studies](#) to better understand risk factors, such as underlying conditions that might put people at increased risk for serious infection. CDC also conducts special studies using hospitalization and treatment data to better understand the clinical course of COVID-19 illness.

## What are the limitations in using case surveillance data to compare jurisdictions or analyze demographic trends?

New COVID-19 cases and deaths are recorded based on data collected and reported by state, local, and territorial health departments. This information can be affected by local testing practices, laboratory capacity, and medical resources. Comparing the COVID-19 situation among jurisdictions should not be based on these rates alone.

When studying the COVID-19 situation in these jurisdictions, the rate of new COVID-19 cases should be combined with other data, including the number of tests performed, the proportion of tests that are positive for SARS-CoV-2, testing policies, excess deaths, and hospital and ICU admission rates.

In addition, jurisdictions vary in the completeness of certain demographic data for COVID-19 cases. Most states have demographic factors like age and sex for most reported cases. However, in many states, the large number of COVID-19 cases has severely strained the ability to report cases with complete demographic information for race and ethnicity. With thousands of cases being reported, completeness of these elements is unlikely to improve in the immediate future for some jurisdictions. Because the racial and ethnic composition of the U.S. population varies by geographic area, comparisons of COVID-19 case information should consider the population of each geographic area. In addition, because completeness of race and ethnicity information may vary by state or geographic area and by other patient factors, such as severity of illness, CDC's case data may not be generalizable to the entire U.S. population.

## How has public health worked to improve reporting of critical information in COVID-19 case surveillance?

CDC has worked with state and jurisdictional health departments to improve reporting of critical case surveillance data elements such as age, race and ethnicity, and death. With thousands of cases being reported, the reporting of some data elements remains low, but state and jurisdictional health departments have continued to make improvements in completeness of data collection for COVID-19 through methods such as automated data flows. As the epidemic changes and number of new cases goes down, CDC and our state, tribal, local, and territorial partners will continue to evaluate the most efficient means to increase the completeness and availability of actionable public health data.

# Surveillance Reports

Does CDC create surveillance reports on COVID-19? 

Yes. On February 12, 2021, we posted the first [COVID Data Tracker Weekly Review](#), which is posted every Friday. This report and newsletter highlight key data from CDC's COVID Data Tracker, including [cases and deaths](#), [variants](#), [testing](#), [hospitalizations](#) and [vaccinations](#). It summarizes important trends in the pandemic and brings together CDC data and reporting in a centralized location.

The COVID Data Tracker Weekly Review replaced the COVIDView report, which was produced weekly from April 3, 2020, through February 5, 2021. An [archive of COVIDView reports](#) is maintained on the CDC web.

COVID-19 surveillance data are also used to produce publications, including CDC's [Morbidity and Mortality Weekly Report \(MMWR\)](#), and to inform [guidance documents](#) to protect people from COVID-19 in a variety of settings.

---

## CDC COVID Data Tracker

What is CDC COVID Data Tracker? 

[CDC COVID Data Tracker](#) is CDC's home for COVID-19 data. It provides surveillance data from across the response, including hospitalizations, vaccinations, demographic information, and daily and cumulative case and death counts reported to CDC since January 21, 2020. Data found on Data Tracker are updated daily. Topics covered on COVID Data Tracker include:

- COVID-19 in Your Community
- Vaccinations
- Cases & Deaths
- Demographic Trends
- Health Care Settings
- Genomic Surveillance
- Testing and Seroprevalence
- People at Increased Risk



## How often is CDC COVID Data Tracker updated?

Tabs on CDC COVID Data Tracker are updated daily unless otherwise specified in the footnote of a given tab. Specifics of data reporting are described in a footnote on each page.

## Is it possible to download data from COVID Data Tracker?

Yes, there are multiple datasets that can be downloaded directly from COVID Data Tracker. To download data from COVID Data Tracker, navigate to the data table in the tab you are viewing and click on the download icon (as seen here).




## How do I download case and death data over time, from January 22, 2020, onward?

To download case and death data over time, including historical data, visit the [U.S. and State Trends tab](#) and click the download icon. Alternatively, the data can be downloaded at [data.cdc.gov](https://data.cdc.gov).

## How do I download the most current aggregate case and death data?


To download the most current aggregate case and death data, visit the [Cases and Deaths by State tab](#) on COVID Data Tracker and click the download icon.


How do I download public use line-level case data from January 1, 2020, onward? 

You can download line-level data, including patient sex, age group, hospitalization status, and race/ethnicity, county, and state of residence (where available) from three available [COVID-19 case surveillance datasets](#).

You can conduct your own analyses using the available datasets to determine the number and selected characteristics of lab-confirmed cases shared with CDC by jurisdictions through a specific date. You can download deidentified CDC case surveillance data, which includes fields for date of first positive specimen collection, case status (lab-confirmed vs. probable), and others. See the next section titled: "CDC Publicly Available Datasets."

## CDC Publicly Available Datasets

Why does CDC provide patient-level datasets to the public? 

Sharing timely and accurate COVID-19 data with the public is a core activity of CDC's COVID-19 Emergency Response as well as a key priority of [CDC's Data Modernization Initiative](#), and the administration's [Executive Order on Ensuring a Data-Driven Response to COVID-19 and Future High-Consequence Public Health Threats](#).  Publicly available datasets are critical for several reasons: open government and transparency, promotion of research, and efficiency (i.e., providing the public, media, and others access to the same data with consistency and supporting information).

## How many datasets are available to the public?

CDC has three COVID-19 case surveillance datasets:

- **COVID-19 Case Surveillance Public Use Data with Geography:** Public use, patient-level dataset with clinical and symptom data, demographics, and state and county of residence. (19 data elements)
- **COVID-19 Case Surveillance Public Use Data:** Public use, patient-level dataset with clinical and symptom data and demographics, with no geographic data. (12 data elements)
- **COVID-19 Case Surveillance Restricted Access Detailed Data:** Restricted access, patient-level dataset with clinical and symptom data, demographics, and state and county of residence. Access requires a registration process and a data use agreement. (32 data elements)

## What privacy protections have been applied to these datasets?

To reduce the risk that these datasets could be used to reidentify persons, CDC designed each dataset accounting for privacy and confidentiality, and conducts ongoing privacy assessments using standard methods and systematically verifies the data prior to release. Strict privacy protections, including data suppression, were applied to all three datasets. See the information included with each dataset for more information.

## Do data in these datasets match the information in the CDC COVID Data Tracker - or on state health department websites?

Although the [CDC COVID Data Tracker](#) and health department websites also report COVID-19 case surveillance data, data may not match the CDC public use datasets due to differences in timing of the creation of the datasets and differences in the timing of reporting and case notification. The three COVID-19 case surveillance datasets are updated monthly, and there is a reporting lag. The CDC COVID-19 Data Tracker is updated daily. When there are differences between numbers of cases reported, data reported by health departments should be considered the most up to date for the state or territory.

Last Updated Apr. 14, 2021  
Content source: [National Center for Immunization and Respiratory Diseases \(NCIRD\), Division of Viral Diseases](#)