**COVID-19 Outcomes by Testing Cohorts: Cases, Hospitalizations, and Deaths User Guide**

Created: June 10, 2020  
Last Updated: June 10, 2020  
Open Data Coordinator: Kemi Eguaze, Department of Health and Mental Hygiene

**What’s In This Dataset?**

This dataset represents COVID-19 outcomes for cohorts of New York City residents based on date of specimen collection for testing for SARS-CoV-2, the virus that causes COVID-19.

The dataset shows outcomes (confirmed cases, hospitalizations, and deaths) for cohorts defined by each date of specimen collection (specimen\_date). For example, if a NYC resident tested positive for SARS-CoV-2 and was subsequently hospitalized, both events would show under the same specimen\_date, indicating the date of specimen collection for the positive test and not the date of the hospitalization. For a comparable dataset showing diagnosis dates for confirmed cases, admission dates for hospitalized patients, and death dates for decedents, see <https://data.cityofnewyork.us/Health/COVID-19-Daily-Counts-of-Cases-Hospitalizations-an/rc75-m7u3>.

Data on laboratory testing and results were passively reported to the NYC Health Department by hospital, commercial, and public health laboratories. In March, April and early May, the NYC Health Department had discouraged people with mild and moderate symptoms from being tested, so our data primarily represent people with more severe illness. Until mid-May, patients with more severe COVID-19 illness were more likely to have been tested and included in these data. Data on hospitalizations for confirmed COVID-19 cases were obtained from direct remote access to electronic health record systems, regional health information organization (RHIOs), and NYC Health and Hospital information, as well as matching to syndromic surveillance. Deaths were confirmed by the New York City Office of the Chief Medical Examiner and the Health Department’s Bureau of Vital Statistics. Data counts are underestimates. Data for patients with specimens recently collected are incomplete. Data are updated daily. For each date of extraction, results for all specimen collection dates are appended to the bottom of the dataset. To analyze the most current data, **only use the latest extract date**. This dataset preserves historical records and source data changes, so each extract date reflects the current copy of the data as of that date. For example, an extract date of 5/1/2020 and extract date of 5/2/2020 will both contain all records as they were as of that extract date. **Without filtering or grouping by extract date, an analysis will almost certainly be miscalculating or counting the same values multiple times**. Lags between specimen collection date and report dates of cases, hospitalizations, and deaths can be assessed by comparing counts for the same specimen collection date across multiple data extract dates.

For further details, visit:

* <https://www1.nyc.gov/site/doh/covid/covid-19-data.page>

**Who Manages This Data?**

Data are assembled by the NYC Department of Health and Mental Hygiene Incident Command System for COVID-19 Response (Surveillance and Epidemiology Branch in collaboration with Public Information Office Branch).

**Get Started With This Data:**

These data can be used to:

* Identify temporal trends in persons tested for SARS-CoV-2 and the number and proportion of persons testing positive.
* Identify temporal trends in the numbers and proportions of hospitalizations and deaths by specimen collection date among persons with confirmed COVID-19.
* Assess lags between specimen collection date and report dates of cases, hospitalizations, and deaths.

**Columns (Fields, Attributes)**

**Extract\_date**: Date of extraction from live disease surveillance database. The extract\_date column can be used to evaluate how data changed over time. To analyze the most current data, **only use the latest extract date**.

**Specimen\_date**: Date of specimen collection, equivalent to diagnosis date. For persons who had an earlier negative or indeterminant test result and a later positive test result, only the specimen collection date corresponding to the first positive test is used.

**Number\_tested**: Count of New York City residents newly tested for SARS-CoV-2. To analyze the most current data, **only use the latest extract date**

**Number\_confirmed:** For patients with specimens collected for SARS-CoV-2 testing on the given specimen\_date, the count of those same patients who were confirmed to be COVID-19 cases. To analyze the most current data, **only use the latest extract date**

**Number\_hospitalized**: For patients with specimens collected for SARS-CoV-2 testing on the given specimen\_date and confirmed to be COVID-19 cases, the count of those same patients who were ever hospitalized. To analyze the most current data, **only use the latest extract date**

**Number\_deaths**: For patients with specimens collected for SARS-CoV-2 testing on the given specimen\_date and confirmed to be COVID-19 cases, the count of those same patients who died. To analyze the most current data, **only use the latest extract date**