

July 2020

Why GAO did this study

The COVID-19 pandemic has resulted in significant loss of life and profoundly disrupted the U.S. economy and society, and the Congress has taken action to support a multifaceted federal response on an unprecedented scale. It is important for decision makers to understand the limitations of COVID-19 data, and the uses and limitations of various methods of analyzing and interpreting those data.

The Coronavirus Aid, Relief, and Economic Security Act (CARES Act) includes a provision for GAO to, in general, conduct monitoring and oversight of the authorities and funding provided to address the COVID-19 pandemic and the effect of the pandemic on the health, economy, and public and private institutions of the U.S. This technology assessment examines (1) collection methods and limitations of COVID-19 surveillance data reported by CDC, (2) approaches for analyzing COVID-19 data, and (3) uses and limitations of forecast modeling for understanding of COVID-19.

In conducting this assessment, GAO obtained publicly available information from CDC and state health departments, among other sources, and reviewed relevant peer reviewed and preprint (non-peer-reviewed) literature, as well as published technical data on specific models.

View [GAO-20-635SP](#). For more information, contact Timothy M. Persons, PhD at (202) 512-6888 or PersonsT@gao.gov, SaraAnn Moessbauer at (202) 512-4943, or MoessbauerS@gao.gov, or Mary Denigan-Macauley, PhD at (202) 512-7114 or DeniganMacauleyM@gao.gov.

COVID-19

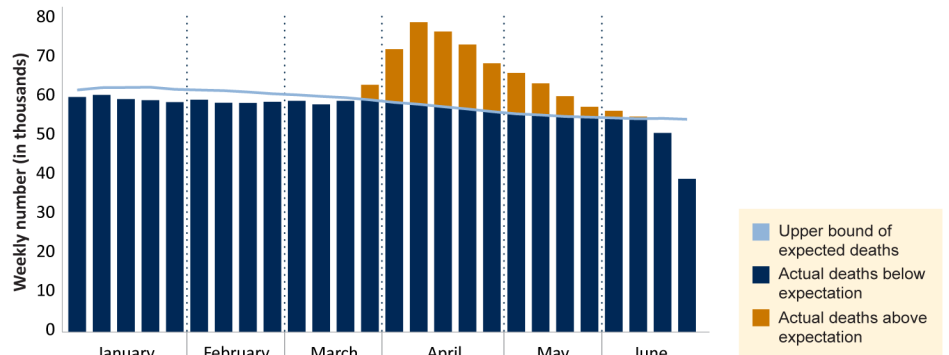
Data Quality and Considerations for Modeling and Analysis

What GAO found

The rapid spread and magnitude of the COVID-19 pandemic have underscored the importance of having quality data, analyses, and models describing the potential trajectory of COVID-19 to help understand the effects of the disease in the U.S. The Centers for Disease Control and Prevention (CDC) is using multiple surveillance systems to collect data on COVID-19 in the U.S. in collaboration with state, local, and academic and other partners. The data from these surveillance systems can be useful for understanding the disease, but decision makers and analysts must understand their limitations in order to interpret them properly. For example, surveillance data on the number of reported COVID-19 cases are incomplete for a number of reasons, and they are an undercount the true number of cases, according to CDC and others.

There are multiple approaches to analyzing COVID-19 data that yield different insights. For example, some approaches can help compare the effects of the disease across population groups. Additional analytical approaches can help to address incomplete and inconsistent reporting of COVID-19 deaths as well. For example, analysts can examine the number of deaths beyond what would normally be expected in the absence of the pandemic. Examining higher-than-expected deaths from all causes helps to address limitations in the reporting of COVID-19 deaths because the number of total deaths is likely more accurate than the numbers of deaths from specific causes. The figure below shows actual deaths from the weeks ending January 1 through June 27, 2020, based on data from CDC’s National Center for Health Statistics, compared with the expected deaths based on prior years’ data. Deaths that exceeded this threshold starting in late March are considered excess deaths that may be related to the COVID-19 pandemic.

Higher-Than-Expected Weekly Mortality for 2020, as of July 14, 2020



Source: GAO analysis of Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS) data. | [GAO-20-635SP](#)

Analysts have used several forecasting models to predict the spread of COVID-19, and understanding these models requires understanding their purpose and limitations. For example, some models attempt to predict the effects of various interventions, whereas other models attempt to forecast the number of cases based on current data. At the beginning of an outbreak, such predictions are less likely to be accurate, but accuracy can improve as the disease becomes better understood.