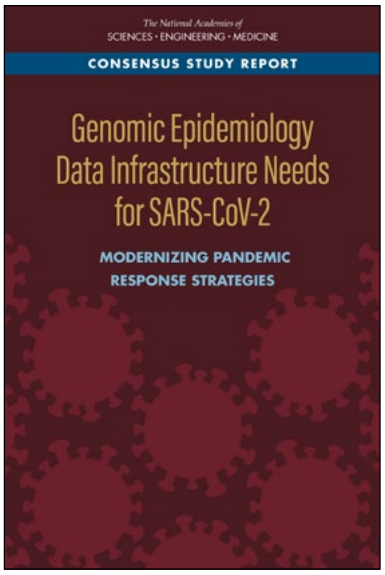
Consensus Study Report

JULY 2020

HIGHLIGHTS

Genomic Epidemiology Data Infrastructure Needs for SARS-CoV-2

Modernizing Pandemic Response Strategies



At the time of writing this report, the virus that causes COVID-19, SARS-CoV-2, has spread worldwide, infecting at least 10 million people and leading to an estimated 500,000 deaths. To best respond to this ongoing threat, policy makers and public health officials need access to cutting-edge pandemic, epidemic, and global preparedness and response strategies. Particularly, the ability to examine the genome of the virus may help public health officials more precisely understand transmission patterns, virus evolution, and potential treatment and prevention interventions. If policy makers and public health officials have access to these data, it may increase the possibility of breaking or delaying virus transmission and in turn reduce illness and deaths from COVID-19.

Recognizing the need to capitalize on these advances during the current COVID-19 pandemic, the Department of Health and Human Services' Office of the Assistant Secretary for Preparedness and Response (HHS/ASPR) and Office of Science and Technology Policy (OSTP) asked the National Academies of Sciences, Engineering, and Medicine to convene an ad hoc committee that would lay out a framework to help define and describe the data needs for a system that can track and correlate viral genome sequences with clinical and epidemiological data. The resulting system, outlined in this report, is designed to help integrate data sharing and data analysis on the evolution of SARS-CoV-2 into public health practice and decision making.

EXISTING DEFICIENCIES IN GENOMIC EPIDEMIOLOGY EFFORTS FOR SARS-COV-2

Several ongoing data projects are using genomic epidemiology to aid the public health response to COVID-19, including federal and nonprofit initiatives. In the United States, the Centers for Disease Control and Prevention's SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance (SPHERES) consortium is coordinating a nationwide genomic sequencing effort. The National Institutes of Health is supporting two initiatives: the National COVID Cohort Collaborative (N3C), a secure portal for patient-level COVID-19 clinical data, and the National Center for Biotechnology Information's Reference Sequence (RefSeq) database. Several regional initiatives have emerged as well, integrating data sharing through existing global efforts like the Global Initiative on Sharing All Influenza Data (GISAID) and Nextstrain.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Due to a number of factors including poor funding, coordination, and capacity, the committee concluded that existing data sources and integration efforts for SARS-CoV-2 genome sequence data are patchy, typically passive, and reactive in the United States. As a result, available data do not represent key population features of individuals who are infected with SARS-CoV-2 and cannot adequately answer many pressing questions about the evolution and transmission of this virus.

To better collect necessary viral sequence data and other associated data, the committee recommended that HHS implement genome sequencing of SARS-CoV-2 on a national scale and ensure that this effort is fully representative of every segment of the population. Particularly, the committee highlighted the need to account for race and ethnicity, gender, age, geography, housing type, displayed symptoms, and transmissibility. To read the full text of this recommendation, see the Recommendations Insert.

BUILDING A FRAMEWORK TO TRACK AND CORRELATE VIRAL GENOME SEQUENCES WITH CLINICAL AND EPIDEMIOLOGICAL DATA

When combined with other types of data, researchers and public health officials may be able to use viral genome sequence data to inform questions related to transmission of a virus, virus evolution, and how the virus manifests as a disease.

However, in order to fully understand these issues, public officials need access to effectively integrated data. Currently, no central repository exists for the collection and curation of such infectious disease outbreak data from federal, state, and local health agencies, health care networks, and public health and clinical laboratories. To address this gap, the committee recommended that HHS develop and invest in a national data infrastructure system that can link genomic, clinical, and epidemiological data. This system should:

- Allow for linkage between these types of data without overburdening laboratories
- Create and foster safe data-sharing practices to protect personally identifying information
- Ensure shared data are standardized, interoperable, flexible, and practical
- Promote effective data collection
- Conduct annual reviews to identify gaps and potential improvements in data collection and sharing

To read the full text of this recommendation, see the Recommendations Insert.

DATA GOVERNANCE AND LEADERSHIP

In the United States, no current federal or state laws protect or mandate the sharing of sequence data from virus samples. Any such data sharing is done voluntarily and generally without concerns about possible regulatory barriers. In contrast, there are federal and state laws protecting clinical and epidemiological data, including the Health Insurance Portability and Accountability Act (HIPAA). Without funding to support the costs associated with gathering and preparing data to be shared with public health authorities and researchers, and without clear regulatory pathways and established infrastructure to support sharing, the sharing of virus sequencing data is suboptimal.

Given the national and interstate threat posed by a pandemic like COVID-19, data sharing needs to happen on a national scale through formalized agreements with federal authorities, rather than purely within state borders or with the federal government on an ad hoc basis. These data sharing and reporting processes should be clearly established and funded prior to any public health emergency. To standardize and encourage the practice of quality data sharing for public health practice during a pandemic, the committee recommended that HHS establish a national leadership structure to facilitate comprehensive data sharing. This structure should include science-driven leadership and governance for the use of SARS-CoV-2 genome sequences. Key components of this structure should also include

- leadership that has the authority and responsibility to identify and prioritize key issues related to data needs;
- a national strategy for using SARS-CoV-2 genome sequences that clearly articulates goals, priorities, and a path for achieving them; and
- a governing board with diverse, relevant expertise that has broad authority to oversee and advise the national strategy for SARS-CoV-2 genome sequencing and the delivery of actionable data for related investigations.

To read the full text of this recommendation, see the Recommendations Insert.

Committee on Data Needs to Monitor the Evolution of SARS-CoV-2

Diane Griffin (*Chair*)
Johns Hopkins Bloomberg School of Public Health

Ralph Baric
University of North Carolina at Chapel Hill

Kent Kester
Sanofi Pasteur

Deven McGraw
Ciitizen Corporation

Alexandra Phelan
Georgetown University Center for Global Health Science and Security

Saskia Popescu
HonorHealth
George Mason University
University of Arizona

Stuart Ray
Johns Hopkins University School of Medicine

David Relman
Stanford University
Palo Alto Health Care System

Julie Segre
National Institutes of Health National Human Genome Research Institute

Mark Smolinski
Ending Pandemics

Paul Turner
Yale University

Deborah Zarin
Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard University

Study Sponsor

U.S. Department of Health and Human Services' Office of the Assistant Secretary for Preparedness and Response

Office of Science and Technology Policy

Study Staff

Lisa Brown
Study Director

Emma Fine
Associate Program Officer

Benjamin Kahn
Associate Program Officer

Steven Moss
Associate Program Officer

Andrew M. Pope
Senior Director, Board on Health Sciences Policy

Liaison to the Standing Committee on Emerging Infectious Diseases and 21st Century Health Threats

Harvey Fineberg
Gordon and Betty Moore Foundation

Chair, Standing Committee on Emerging Infectious Diseases and 21st Century Health Threats

To read the full report, please visit
nationalacademies.org/SARS-CoV-2-Data-Infrastructure-Needs

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.
www.nationalacademies.org