United States Government Accountability Office

Testimony

Before the Subcommittee on Oversight and Investigations, Committee on Energy and Commerce, House of Representatives

For Release on Delivery
Expected at 2:00 p.m. ET
Wednesday, February 1, 2023

PANDEMIC ORIGINS

Technologies, Challenges, and Policy Options to Support Investigations

Accessible Version

Statement of Dr. Karen L. Howard,
Acting Chief Scientist and Director,
Science, Technology Assessment, and Analytics

Chair Griffith, Ranking Member Castor, and Members of the Subcommittee

Thank you for the opportunity to discuss our work on pandemic origins. My testimony today summarizes our January 2023 technology assessment entitled *Pandemic Origins: Technologies and Challenges for Biological Investigations*.[1] GAO's technology assessments focus on examining technologies and identifying their challenges and benefits. The report I am discussing today examines technologies—including tools and methods—used to investigate the origin of infectious diseases that lead to pandemics.[2]

Given the magnitude of the health and economic costs of pandemics, there is a need to better understand how and where they originate.[3] According to scientific literature, most pandemics where the origin is known were caused by the natural transmission of a pathogen through animal-to-human contact, also known as zoonotic transmission. A pandemic could also potentially be initiated through the accidental infection of an individual or individuals by a pathogen in a laboratory setting, or infections outside the laboratory caused by an accidental or intentional release of the pathogen from a laboratory. For example, research suggests the 1977-1978 H1N1 influenza pandemic may have been the result of a laboratory accident or other cause.[4] Determining the likely origin of pandemics is challenging and requires information

[1]GAO, *Pandemic Origins: Technologies and Challenges for Biological Investigations*, GAO-23-105406 (Washington, D.C.: January 27, 2023).

[2]Determination of a pandemic's origin has some level of inherent scientific uncertainty. For our January 2023 report and this testimony statement, we use the term "origin" to mean "likely origin," acknowledging this uncertainty.

[3]As of the week ending January 7, 2023, the U.S. had about 1,090,000 reported deaths attributed to COVID-19. A recent assessment estimated the human and economic cost of the COVID-19 pandemic to the U.S. totaled more than $10 trillion. The 2009 H1N1 influenza pandemic resulted in approximately 61 million cases and 12,500 deaths in the U.S. Prior to a successful vaccination campaign that eradicated smallpox in 1980, the disease killed approximately 300 million people globally between 1900 and 1980.

[4]Other causes suggested for the 1977-1978 H1N1 influenza pandemic include deliberate release of the virus or a vaccine trial mishap. See M. Rozo and G.K. Gronvall, "The Reemergent 1977 H1N1 Strain and the Gain-of-Function Debate," *mBio*, vol. 6 (2015):e01013-15.

gathered from established methods for disease outbreak investigations that may, in some cases, take a decade or longer of research to acquire.[5]

Our January 2023 report and my statement today address key technologies available for pandemic origin investigations; strengths and limitations of these technologies; and cross-cutting challenges researchers face in trying to determine a pandemic's origin.[6]

To understand the available technologies and challenges in determining the origins of pandemics, we convened a 3-day meeting of 27 experts in March 2022 with assistance from the National Academies of Sciences, Engineering, and Medicine.[7] We also examined peer-reviewed scientific literature and other documents, including the 2022 National Biodefense Strategy and reports from the World Health Organization, Department of Health and Human Services' (HHS) Centers for Disease Control and Prevention (CDC), Office of the Director of National Intelligence, the Johns Hopkins Center for Health Security, and select national laboratories. Further, we interviewed officials and researchers from 11 relevant federal agencies as well as nonfederal experts with a diverse set of perspectives on the science and application of these technologies. Additional information about our scope and methodology can be found in our January 2023 report. We performed the work on which this testimony is based in accordance with all sections of GAO's Quality Assurance Framework that are relevant to technology assessments.

Technologies Are Mature and Can Help Inform Pandemic Origin Investigations

Several key technologies and approaches can help inform investigations of a pandemic's origin, including: genetic sequence analysis; pathogen exposure monitoring and disease tracking; and laboratory-based

[5]For example, it took approximately 13 years to determine the origin of the SARS-associated coronavirus (SARS-CoV) pathogen that caused the 2002-2003 SARS pandemic. While the first human outbreak of H1N1 occurred in Mexico in early 2009, it wasn't until 2016 that it was established that the virus jumped from pigs to humans in central Mexico. The origin of the Ebola virus remains inconclusive.

[6]For the purposes of our report, the term "technologies" includes the instruments, techniques, skills, methods, and processes used in pathogen characterization.

[7]Meeting participants were from academia, business, and nonprofit organizations. For a complete list of participants, see Appendix I of this statement.

pathogen studies. However, to effectively apply these technologies, researchers require samples and data obtained from infected people, animals, and the environment in or around outbreak areas from as early in an outbreak as possible.

Genetic sequence analysis. Experts told us that they consider genomic sequencing one of the key technologies for pandemic origin investigations due to its speed, accuracy, and cost. Genomic sequencing allows researchers to generate a pathogen's genetic sequence. This genetic sequence is then analyzed using bioinformatics tools and compared to reference genetic sequences stored in databases to identify matches with other known pathogens, mutations in the sequences, potential genetically-engineered sequences, and likely relationships to the nearest relatives. For example, researchers used genetic sequence analysis to help establish the likely natural origins of the 2002-2003 SARS pandemic, the 2009 H1N1 influenza pandemic, and the initial MERS outbreak in 2012. However, some laboratory-based genetic modifications may be indistinguishable from natural variations. For example, some traditional genetic engineering techniques and newer genome editing tools—such as CRISPR—may not leave readily detectable traces of genetic modification. Further, repeated growth of the pathogen in laboratory animals or cell cultures may result in changes (i.e., mutations) in the pathogen that closely mimic the natural processes of evolution.

Pathogen exposure monitoring and disease tracking. Technologies such as serology (i.e., blood analysis) and epidemiological surveillance—tracking a disease as it moves through a population—are also key technologies for pandemic origin investigations. These technologies allow researchers to monitor pathogen infection and disease occurrence in human and animal populations. For example, serology surveillance in people and camels provided two key pieces of information that contributed to the determination that camels were direct sources of human infection with MERS-CoV.

Laboratory-based pathogen studies. The exact processes by which some pathogens adapt to infect and transmit between humans are not well-understood, which may limit investigators' abilities to establish the origin of a pandemic. Therefore, laboratory-based pathogen studies using cell cultures or animals may provide evidence supporting known natural or unusual patterns of spread. The latter may indicate a possible laboratory-related origin. For example, researchers studying pandemic H1N1 influenza virus in ferrets identified the viral genes, proteins of transmission, and host receptor sites that drive different routes of

transmission. The results of these studies supported the conclusion that this virus likely originated from animal-to-human transmission. However, results from controlled laboratory studies may not accurately represent the natural environment, making it difficult for researchers to clearly distinguish between natural versus laboratory-controlled patterns of spread.

Investigators need access to samples and data, particularly from infected or exposed individuals, from as early in an outbreak and as geographically close to the first reported human disease cases as possible, for these technologies to be effective in determining a pandemic's origin. However, certain countries may refuse or limit researchers' access to field sites, facilities, data, or people. For example, researchers and agency analysts reported that uncertainty still exists about where the first SARS-CoV-2 infections occurred because of a lack of clinical samples available for serological and genetic analyses as well as a lack of epidemiological data from the earliest cases.

Cross-Cutting Challenges that Hinder Pandemic Origin Investigations and Policy Options that May Help Address Them

According to experts, technologies are not the limiting factor for determining the likely origin of a pandemic. Experts identified three cross-cutting key challenges that hinder researchers trying to investigate the origin of a pandemic:

- Lack of sufficient access to samples and genetic sequence data;
- Lack of standardized processes for submitting, accessing, and using genetic sequence data stored in databases around the world; and
- Lack of a sufficient and skilled interdisciplinary workforce.

We identified five policy options that may help address these challenges and help improve the ability of researchers to respond more quickly and effectively to future pandemics.

Challenge: Lack of sufficient access to samples and genetic sequence data. Privacy concerns, general mistrust, perceived infringements on a country's sovereignty, or fear of negative consequences may limit access to samples and data. Further, even if

researchers have access to samples and data, their ability to extract suitable information may be limited by a lack of standardized processes. For example, health officials may collect samples for a purpose other than pathogen surveillance, or store and process the data obtained from the samples in a way incompatible with what is needed for effective investigations. Additionally, no one entity is responsible for determining and enforcing standardized processes.

Policy Option: Experts and some agency officials told us that federal policymakers, such as the Department of State, and others could help address this challenge in advance of future outbreaks by establishing comprehensive multilateral, international agreements for accessing and sharing genetic sequence samples and data. These proactive agreements could include definitions of the roles and responsibilities of international investigation teams and incentives for adherence, helping ensure more timely access to critical information. Negotiating or modifying agreements each time a pandemic occurs is not effective because of the speed with which pandemics spread. That is, agencies do not have months to negotiate a series of bilateral agreements with every country every time an outbreak occurs. Instead, policymakers and others could proactively:

- Develop multilateral sample and data-sharing agreements—for example, agreements which include expectations of timely access to samples and detailed standards for sample collection;
- Work with international health organizations, such as the World Health Organization, to identify and address barriers to establishing multilateral, international agreements for ensuring access to genetic sequence samples and data, and support the development of such agreements; and
- Seek agreement with stakeholders on incentives for participation, such as equitable access to vaccines and therapeutics. These incentives could also include economic assistance and assurances to mitigate stigmatization when promptly sharing samples and genetic sequence data.

A key benefit of establishing these proactive agreements is ensuring timely access to genetic information and samples in the critical beginning stages of a pandemic and throughout an origin investigation. Such access may help in the determination of a pandemic's origin. However, some countries may be unwilling to participate in these agreements because of concerns related to national sovereignty, among other reasons. Further,

identifying an appropriate responsible entity to determine and monitor whether countries are following agreed-upon standard processes may be challenging.

Challenge: Lack of standardized processes for genetic sequence databases prevents researchers from analyzing data effectively. To investigate the origin of a pandemic, researchers need access to genetic sequence data, which may be stored in multiple databases.[8] Experts cited three main issues with working across multiple databases:

- Each genetic sequence database may have different processes for submitting, accessing, and using the data. As a result, gathering all of the data necessary to investigate the origin of a pandemic can be challenging.
- Genetic sequence databases generally lack standardized user interfaces for data submission and access, and some existing user interfaces can be cumbersome. The need for different procedures to submit and retrieve data from relevant databases can be time-consuming and inefficient for researchers.
- Metadata such as the date and location of sample collection are crucial for investigating the origin of a pathogen, but their availability and quality may vary. For example, although GenBank® allows users to report specific locations where samples were collected, a 2017 study estimated that 99 percent of records do not include that information. If the information does exist, researchers may still have to perform additional steps of integrating this information from other fields in the sample's record, which is challenging and may affect the reliability of the location data.

Rapid growth of big data

A 2015 study predicted that, by 2025, genomics research worldwide will generate between 2 and 40 exabytes of data annually. (For reference, 1 exabyte equals 1 billion gigabytes.) This would make genomics one of the most challenging domains of Big Data in terms of data acquisition, storage, distribution, and analysis.

Accommodating the expected growth of genomic data will require advancements in

[8]These databases include GenBank®, Global Initiative on Sharing All Influenza Data (GISAID), and European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI).

computational speed and power, as well as algorithms optimized for Big Data.

Source: GAO review of literature. | GAO-23-106562

These issues may be exacerbated by the immense scale and continued growth of genetic sequence data. (See text box for a prediction on the future growth of genomic data.)

As the amount of data in each database grows, and as more databases are added, standardized processes are crucial to ensure that researchers can compile, analyze, and share all the genetic sequence data necessary to investigate the origin of a pandemic. However, it is unclear whether the existing infrastructure of the independent databases worldwide can support the growth of genomic data.

Policy Options: Experts identified two possible options policymakers could consider to address this challenge of a lack of standardized processes for genetic sequence databases. First, federal policymakers and others—such as HHS, current database providers, developers, and users—could collaborate to identify and develop standardized processes for submission of and access to data in databases such as GenBank to support pandemic origin investigations. Second, policymakers could encourage the improvement of current, or development of new, genetic sequence database tools—such as user interfaces or application programming interfaces (API)—of current databases, or incentivize the creation of new user interfaces or APIs to help investigators determine a pandemic's origin more effectively.[9]

The key benefits of developing standardized processes and improving interfaces for database use include ensuring the consistency and quality of submitted data to help researchers access and compare genetic sequences and address the projected future growth in genetic sequence data. However, standardized processes and interfaces may be difficult and expensive to develop, and it may be challenging for multiple stakeholders to agree on what data and interface features are important.

Challenge: The global research community lacks a sufficient and skilled interdisciplinary workforce. Pandemic origin investigations require a highly skilled workforce with expertise in multiple fields. We identified four main reasons it can be hard to develop and retain such a workforce:

- Demand for workers in relevant fields tends to increase when pandemics occur and decrease when pandemics end. Likewise,

[9]An application programming interface (API) enables machine-to-machine communication, allowing users to obtain real-time data updates.

funding for relevant research tends to fluctuate. This makes it challenging to keep the workforce in readiness (i.e., available and proficient) to conduct investigations promptly when pandemics occur.

- Pandemic origin investigations require expertise in multiple fields such as biology, virology, microbiology, immunology, epidemiology, ecology, genomics, bioinformatics, and computer science. However, experts we interviewed told us the current workforce is siloed because of academic structures, funding priorities, and grant processes. This makes it challenging to build and maintain the multidisciplinary workforce necessary to conduct investigations.
- The current uneven global distribution of the workforce leads to political and logistical challenges during a pandemic. For example, a 2021 study of one country concluded that inadequate sequencing capacity because of limited skillsets, among other factors, hindered biosurveillance during the COVID-19 pandemic.[10]
- Some researchers told us that they faced criticism because of their involvement in investigating the origin of a pandemic, particularly when their conclusions were considered controversial. These researchers said they and others may be reluctant to participate in further investigations because of personal and professional risks.

Policy Option: To address this challenge, policymakers could incentivize the development, retention, and growth of a workforce—including in areas considered hot spots of emerging infectious disease—with the critical skills to conduct or support the work of characterizing the likely origin of a pandemic. One way to implement this policy option is by creating international partnerships, among other things, and leveraging or creating training programs to encourage workforce growth and retention.

A sufficient and skilled workforce would ensure that the workforce is not concentrated in any geographic region. A trained workforce skilled in origin investigations could also contribute to other areas such as public health, or other related activities. However, the scientific community may resist alteration to current academic structures, and it may be challenging to adapt priorities, processes, and funding in a sufficiently timely manner needed to respond to a pandemic. As a result, attracting qualified people

[10]M. Dzobo et al., “Inadequate SARS-CoV-2 Genetic Sequencing Capacity in Zimbabwe: A Call to Urgently Address this Key Gap to Control Current and Future Waves,” *IJID Regions*, vol. 1 (2021): ep. 3-4. https://doi.org/10.1016/j.ijregi.2021.09.004.

into the necessary workforce fields may be challenging if those fields are marginalized and underfunded.

Cross-Cutting Policy Option: Develop a national pandemic origin strategy. While the first four policy options may help address the specific challenges we identified and help improve the ability of researchers to respond more quickly and effectively to future pandemics, we found that a national strategy could help to address all of these challenges. For example, the 2022 National Biodefense Strategy and Implementation Plan includes an Early Warning priority area that encompasses targets and corresponding actions related to determining the origin of biological events, including infectious disease outbreaks. However, augmenting the 2022 Strategy or developing a separate strategy with more specifics, such as specifying how the lead and support departments and agencies will coordinate and collaborate, could better position the nation to play a leading role in pandemic origin investigations. For example,

- Federal policymakers could augment the 2022 National Biodefense Strategy to specify how lead and support departments and agencies will coordinate and collaborate with domestic and international partners to address pandemic origin investigations; or
- Federal policymakers could develop a new, standalone, national strategy focused on pandemic origin investigations that describes how federal entities will coordinate and collaborate with domestic and international partners on such investigations.

The key benefits of a national strategy with federal coordination and collaboration leadership include increasing preparedness for future pandemic origin investigations and mitigating health and economic costs. However, allocating resources and defining how federal agencies and others will collaborate may be challenging because of the number and types of entities with relevant expertise. Further, during nonpandemic periods, other priorities and needs may arise and make it challenging to provide sustained resources and support needed for maintaining a national strategy.

In closing, we found that technologies are mature and available for helping inform the origin of pandemics, but several non-technological challenges hinder such investigations. To address these challenges, we proposed five policy options for consideration. These options would better position our nation to deal with future pandemics, in particular, by crafting

multilateral agreements for sample and data sharing and developing a targeted national strategy for pandemic origin investigations.

Chair Griffith, Ranking Member Castor, and Members of the Subcommittee, this concludes my statement. I would be pleased to respond to any questions you or other Members may have.

GAO Contact and Staff Acknowledgments

If you or your staff have any questions about this testimony, please contact Karen L. Howard at (202) 512-6888 or howardk@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this statement. Key contributors to this testimony include Hayden Huang (Assistant Director), Michael Dickens (Analyst-in-Charge), Calaera Powroznik, Craig Starger, and Adam Wells. Additional contributors to the prior work on which this testimony is based are listed in our January 2023 report.

Appendix I: Expert Meeting Participants

For the report on which this testimony is based, we convened a 3-day meeting of 27 experts with assistance from the National Academies of Sciences, Engineering, and Medicine to inform our work on technologies for determining pandemic origin; the meeting was held virtually March 22–24, 2022. The experts who participated in this meeting are listed below.

David B. Allison, PhD; Dean, Distinguished Professor and Provost Professor, Indiana University–Bloomington School of Public Health

Jesse Bloom, PhD; Professor, Basic Sciences Division; Professor, Herbold Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center

Roger Brent, PhD; Professor, Basic Sciences Division; Professor, Public Health Sciences Division, Fred Hutchinson Cancer Research Center

James Diggans, PhD; Distinguished Scientist, Bioinformatics and Biosecurity, Twist Bioscience

Joshua Dunn, PhD; Head of Design, Ginkgo Bioworks, Inc.

Livia Schiavinato Eberlin, PhD; Associate Professor, Department of Surgery, Baylor College of Medicine

Patrick Fitch, PhD; Associate Director of Chemical, Earth and Life Sciences, Los Alamos National Laboratory

A. Oveta Fuller, PhD; Associate Professor of Microbiology and Immunology, Medical School at University of Michigan

Gigi Kwik Gronvall, PhD; Senior Scholar, Johns Hopkins Center for Health Security; Associate Professor, Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health

India Hook-Barnard, PhD; Executive Director, Engineering Biology Research Consortium

Katrina Kalantar, PhD; Computational Biology Lead, Infectious Diseases, Chan Zuckerberg Initiative

Ali S. Khan, MD, MPH, MBA; Dean, College of Public Health, University of Nebraska Medical Center (UNMC); Former Assistant Surgeon General, U.S. Public Health Service

Andy Kilianski, PhD; Senior Director for Emerging Infectious Diseases, International AIDS Vaccine Initiative (IAVI); Adjunct Professor, Schar School of Policy and Government, George Mason University

Sergios-Orestis Kolokotronis, PhD, MPhil, MA; Assistant Professor, Department of Epidemiology and Biostatistics, School of Public Health, The State University of New York (SUNY) Downstate Health Sciences University

Suresh Kuchipudi, BVSc, MVSc, PhD, PGCHE, FHEA, Dip. ACVM, MBA; Professor and Endowed Chair in Emerging Infectious Diseases, Pennsylvania State University; Associate Director, Penn State Animal Diagnostic Laboratory (ADL)

Jacob Lemieux, MD, DPhil; NIH-funded Physician/Scientist, Division of Infectious Disease, Massachusetts General Hospital (MGH) and Harvard Medical School (HMS)

Bronwyn MacInnis, PhD; Director of Pathogen Genomic Surveillance, Infectious Disease and Microbiome Program, Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard

Alemka Markotić, MD, PhD; Director, University Hospital for Infectious Diseases, Zagreb, Croatia; Head of Department for Research and Head of Clinical Department for Urinary Tract Infections and Full Professor, Medical School, University of Rijeka and Catholic University Zagreb; Associate Member, Croatian Academy

Jonna Mazet, DVM, MPVM, PhD; Vice Provost – Grand Challenges, University of California (UC) Davis; Chancellor's Leadership Professor of Epidemiology and Disease Ecology and Founder, One Health Institute, UC Davis School of Veterinary Medicine

Folker Meyer, PhD; Professor of Data Science, University Hospital, University of Duisburg-Essen

Tara O'Toole, MD, MPH; Senior Fellow, In-Q-Tel; Director, IQT Lab, BiologyNext

Rushika Perera, PhD; Associate Professor, Department of Anatomy, University of California (UC) San Francisco

Brian Plew; Director, Public Health Solutions, Thermo Fisher Scientific

David Relman, MD; Thomas C. and Joan M. Merigan Professor in Medicine, Professor of Microbiology & Immunology, Senior Fellow, Center for International Security and Cooperation, Stanford University; Chief of Infectious Diseases, Veterans Affairs Palo Alto Health Care System

Aaron Streets, PhD; Assistant Professor in Bioengineering, University of California (UC) Berkeley; Core Member, Biophysics Program and Center for Computational Biology, Investigator, Chan Zuckerberg Biohub

David Walt, PhD; Hansjörg Wyss Professor of Bioinspired Engineering, Harvard Medical School; Professor of Pathology, Brigham and Women's Hospital; Core Faculty Member, Wyss Institute at Harvard University

Susan Weiss, PhD; Professor and Vice Chair, Department of Microbiology and Co-Director, Penn Center for Research on Coronaviruses and Other Emerging Pathogens, Perelman School of Medicine, University of Pennsylvania; Governor, American Academy of Microbiology

GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through our website. Each weekday afternoon, GAO posts on its website newly released reports, testimony, and correspondence. You can also subscribe to GAO's email updates to receive notification of newly posted products.

Order by Phone

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's website, https://www.gao.gov/ordering.htm.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

Connect with GAO

Connect with GAO on Facebook, Flickr, Twitter, and YouTube.
Subscribe to our RSS Feeds or Email Updates. Listen to our Podcasts.
Visit GAO on the web at https://www.gao.gov.

To Report Fraud, Waste, and Abuse in Federal Programs

Contact FraudNet:

Website: https://www.gao.gov/about/what-gao-does/fraudnet

Automated answering system: (800) 424-5454 or (202) 512-7700

Congressional Relations

A. Nicole Clowers, Managing Director, ClowersA@gao.gov, (202) 512-4400, U.S. Government Accountability Office, 441 G Street NW, Room 7125, Washington, DC 20548

Public Affairs

Chuck Young, Managing Director, youngc1@gao.gov, (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, DC 20548

Strategic Planning and External Liaison

Stephen J. Sanford, Managing Director, spel@gao.gov, (202) 512-4707
U.S. Government Accountability Office, 441 G Street NW, Room 7814,
Washington, DC 20548