TOWARDS EPIDEMIC PREDICTION: FEDERAL EFFORTS AND OPPORTUNITIES IN OUTBREAK MODELING

PRODUCT OF THE
Pandemic Prediction and Forecasting
Science and Technology Working Group
OF THE NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

EXECUTIVE OFFICE OF THE PRESIDENT OF THE UNITED STATES

December 2016

EXECUTIVE OFFICE OF THE PRESIDENT
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL
WASHINGTON, D.C. 20502

Dear Colleagues:

Infectious diseases threaten public health, national security, and economic prosperity. Outbreaks are spreading rapidly in today's increasingly interconnected world. New diseases emerge as pathogens spill over from animals to humans, while previously treatable pathogens grow more dangerous as they become resistant to drugs. The recent outbreaks of Ebola in West Africa and Zika across the Americas highlight the need for global infectious-disease preparedness.

An emerging science offers new opportunities for an anticipatory, data-driven approach to predicting and projecting infectious disease outbreaks. Building on advances in fields such as genomic technology, pathogen and host biology, bioinformatics, ecology, and machine learning, scientists are developing computational tools for predicting the occurrence of an outbreak and projecting its course. Public health responders increasingly use these tools to determine how best to contain an outbreak. In the future, understanding the processes that drive disease emergence and transmission could help to predict and prevent large-scale outbreaks.

The enclosed report aims to advance the science and application of outbreak prediction. It highlights Federal investments in modeling of human, animal, and plant disease outbreaks, including foundational research into disease emergence, predictive modeling contests, and development of decision-support technologies for outbreak response. The report identifies key challenges and recommends Federal actions to surmount them.

Chartered in April 2014, the NSTC Pandemic Prediction and Forecasting Science and Technology Working Group coordinates Federal initiatives to accelerate the development of outbreak prediction capabilities. The Working Group drafted the report based on its assessment of experiences across and beyond the Federal government in outbreak modeling. The Working Group also held workshops and seminars to engage researchers and stakeholders from universities, industry, public health associations, and other non-Federal organizations.

The science and technology opportunities identified in this report will not eliminate uncertainties in modeling outbreak dynamics. Public health surveillance and planning must remain the mainstays of outbreak preparedness. The recommended steps promise to enhance these measures and reduce the impact of outbreaks through more timely and appropriate public health decisions.

Sincerely,

Jean-Paul Chretien
Chair, Pandemic Prediction and Forecasting
Science and Technology Working Group
National Science and Technology Council

About the National Science and Technology Council

The National Science and Technology Council (NSTC) is the principal means by which the Executive Branch coordinates science and technology policy across the diverse entities that make up the Federal research and development (R&D) enterprise. One of the NSTC's primary objectives is establishing clear national goals for Federal science and technology investments. The NSTC prepares R&D packages aimed at accomplishing multiple national goals. The NSTC's work is organized under five committees: Environment, Natural Resources, and Sustainability; Homeland and National Security; Science, Technology, Engineering, and Mathematics (STEM) Education; Science; and Technology. Each of these committees oversees subcommittees and working groups that are focused on different aspects of science and technology. More information is available at www.whitehouse.gov/ostp/nstc.

About the Office of Science and Technology Policy

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization, and Priorities Act of 1976. OSTP's responsibilities include advising the President in policy formulation and budget development on questions in which science and technology are important elements; articulating the President's science and technology policy and programs; and fostering strong partnerships among Federal, state, and local governments, and the scientific communities in industry and academia. The Director of OSTP also serves as Assistant to the President for Science and Technology and manages the NSTC. More information is available at www.whitehouse.gov/ostp.

About the Pandemic Prediction and Forecasting Science and Technology Working Group (PPFST WG)

The PPFST WG provides Federal departments and agencies a forum to coordinate and advance research, technological development, and practice in the prediction of infectious disease outbreaks in humans, animals, or plants to minimize their adverse health, economic, and security impact.

Acknowledgments

Special thanks to Koya Allen, Jason Asher, Manju Bangalore, Matthew Biggerstaff, Emily Briskin, Russ Bulluck, Eleanor Celeste, Kelly Chafin, Ruvani Chandrasekera, Anne Cheever, Andrew Clements, Paul Cross, Helen Cui, Sebastien Darteville, Matthew Davenport, Amy Delgado, Alina Desphande, Thomas Doker, Aaron Firoved, Kimberly Forde-Folle, Glenn Fowler, Robert Garrett, Cyril Gay, Dylan George, Misrak Gezmu, Samantha Gibbs, Ardath Grills, John Hannan, Camille Harris, Michelle Hawkins, John Haynes, Matthew Hepburn, Allison Hinds, Robert Huffman, Kristen Jordan, Emily Kahn, Richard Kiang, Christopher Kiley, Adrienne Keen, Sunil Kumar, Paul Lewis, Michael Johansson, Franca Jones, Tracy Laporte, Nancy Lin, Kenneth Linthicum, Hannah Lobel, Martin Meltzer, Ronald Meris, Jeff Morgan, Suzan Murray, Phil Newton, Anthony Nguy-Robertson, Kelly Patyk, Lisa Pearse, Diann Prosser, Michael Prouty, Teresa Quitugua, Veerasamy Ravichandran, Columb Rigney, Gary Roselle, Bruce Ruscio, Caitlin Rivers, Alison St. John, Michael Sardelis, Julie Schafer, Sam Scheiner, Tricia Schmitt, David Shepherd, Ian Simon, Lynn Slepski, Tom Slezak, Andrew Stern, Mark Teachman, Sally Tinkle, Juli Trtanj, Cecile Viboud, Daniel Walsh, Melissa Warden, Pai-Yei Whung, and Jody Wireman.

Copyright Information

Report prepared by

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL
COMMITTEE ON HOMELAND AND NATIONAL SECURITY
SUBCOMMITTEE ON BIOLOGICAL DEFENSE RESEARCH AND DEVELOPMENT
PANDEMIC PREDICTION AND FORECASTING SCIENCE AND TECHNOLOGY WORKING GROUP

National Science and Technology Council

Chair
John P. Holdren
Assistant to the President for Science
and Technology and Director,
Office of Science and Technology Policy

Staff
Afua Bruce
Executive Director
National Science and Technology Council

Committee on Homeland and National Security

Chairs
Steve Fetter
Associate Director for National Security and
International Affairs
Office of Science and Technology Policy

Reggie Brothers
Under Secretary for Science and Technology
Department of Homeland Security

Stephen Welby
Principal Deputy, Assistant Secretary of Defense
Research Engineering
Department of Defense

Staff
Tod Campanion
Acting Director, Interagency Office, S&T Directorate
Department of Homeland Security

Subcommittee on Biological Defense Research and Development

Chairs
Gerald Epstein
Assistant Director for Biosecurity and Emerging
Technologies
Office of Science and Technology Policy

Staff
Andrew Rickles
Senior Biodefense Advisor
Department of Homeland Security

Pandemic Prediction and Forecasting Science and Technology Working Group

Chairs
Jean-Paul Chretien
Senior Policy Advisor, Biological Threat Defense
Office of Science and Technology Policy

Michael Johansson
Biologist, Dengue Branch
Centers for Disease Control and Prevention

Paul Lewis
Chief, Innovation & Evaluation
Armed Forces Health Surveillance Branch,
Defense Health Agency

Staff
Robert Huffman
Deputy, Biosurveillance Strategy & Policy
Joint Program Executive Office for Chemical and
Biological Defense, Department of Defense

Members

Department of Agriculture
Department of Commerce
Department of Defense
Department of Energy
Department of Health and Human Services
Department of Homeland Security
Department of Interior
Department of Justice
Department of State
Department of Transportation
Department of Veterans Affairs
Environmental Protection Agency
National Aeronautics and Space Administration
National Science Foundation
Office of the Director of National Intelligence
Smithsonian Institution
United States Agency for International Development

Office of Management and Budget
Office of Science and Technology Policy
National Security Council

Table of Contents

Executive Summary 1

Introduction 3

  The Threat of Emerging Infectious Disease Outbreaks 3

  Infectious Disease Outbreak Modeling 4

  Federal Efforts in Outbreak Modeling 7

Challenges and Recommendations 10

  Urgency of Action 15

Appendix 1 17

Appendix 2 19

References 21

Executive Summary

Infectious disease outbreaks threaten global health, economic vitality, and U.S. national security. Infectious disease emergencies prompting U.S. and international response efforts have involved previously unknown pathogens, such as Middle East Respiratory Syndrome Coronavirus and Severe Acute Respiratory Syndrome Coronavirus, as well as known pathogens causing outbreaks of unprecedented magnitude, such as Zika virus and Ebola virus. These and many other emerging or re-emerging infectious diseases of humans are zoonotic—they originate, and may persist, in non-human animal species. Other pathogens affecting only animals or plants, such as foot-and-mouth disease and sudden oak death, respectively, may threaten food security, trade, and ecosystems. A convergence of factors including globalized travel and trade, climate change, urbanization, and agricultural practices contributes to infectious disease outbreaks of humans, animals, and plants. In recent years, a new, interdisciplinary science has drawn on advances in pathogen biology, genomics, bioinformatics, and machine learning to develop computer-based models to predict infectious disease outbreaks, or project their course. These computational tools increasingly support a range of decisions in outbreak preparedness and response, and were used extensively in recent outbreaks. U.S. public health preparedness strategies recognize the potential of these tools to improve outbreak preparedness and response, and many Federal agencies fund research and development programs in outbreak prediction and modeling.

Drawing on this broad interagency experience, the Pandemic Prediction and Forecasting Science and Technology Working Group (PPFST WG), of the National Science and Technology Council, in this report identifies challenges in outbreak prediction and modeling and offers recommendations for Federal actions to advance the development and effective application of outbreak prediction capabilities. The three major challenge areas, and recommendations to address the challenges (to be implemented within existing authorities and resources), are:

<u>Challenge #1: Data- and information-sharing.</u>

- *Recommendation 1*: Identify key questions likely to arise during outbreak response decision-making to help define and prioritize data collection and modeling goals.
- *Recommendation 2*: Adopt a common plan for identifying the minimal essential, publicly-releasable data to support outbreak decision-making and for rapidly processing, updating, and sharing these data publicly in a form readily usable for analysis by computational tools.
- *Recommendation 3*: Collaborate with professional organizations, state and local public health offices, and other stakeholders to develop standards for collecting and sharing the types of data needed for modeling and analysis during outbreaks.
- *Recommendation 4*: Ensure that data obtained with Federal funds that may address key questions during outbreak response decision-making is shared rapidly with other Federal stakeholders and in a form readily usable for analysis with computational tools (and with meta-data that describes potential restrictions).
- *Recommendation 5*: Ensure that analytical products (e.g., scientific manuscripts) generated with Federal funds that may address key questions during outbreak response decision-

making, or inform remediation efforts following such an emergency, and that are appropriate for public release, are shared rapidly with the public free of charge.

- *Recommendation 6*: Develop science and technology solutions to the privacy, security, and related challenges that impede data access, such as de-identification algorithms and synthetic datasets.

Challenge #2: Outbreak model development and decision support.

- *Recommendation 7*: Enable Federal efforts to systematically identify outbreak modeling studies as they become available; conduct controlled comparisons of models if source code is accessible; assess the quality of models and modeling studies; and maintain current estimates of key epidemiological and pathogen parameters through synthesis of modeling results.
- *Recommendation 8*: Strengthen communication and coordination between modeling groups and Federal offices requiring outbreak decision support.
- *Recommendation 9*: Strengthen support for development of standards or best practices to assess outbreak model attributes and performance.
- *Recommendation 10*: Strengthen support for controlled comparisons of outbreak modeling approaches to identify preferred approaches for possible real-world scenarios.
- *Recommendation 11*: Strengthen support for development of decision-support methods and tools to facilitate synthesis and adjudication of results across multiple models.
- *Recommendation 12*: Establish criteria for public release of outbreak model code developed with Federal funds, and ensure that code that is appropriate for public release is shared publicly, free of charge, no later than the time that the scientific paper describing the model is published.
- *Recommendation 13*: Establish active, population-based surveillance testbeds for rigorous development, comparison, and validation of infectious disease models (first consideration should be given to building modeling components into existing epidemiological surveillance or research efforts).

Challenge #3: Science of disease emergence.

- *Recommendation 14*: Develop a "One Health" national strategy for predictive modeling of potential infectious disease threats globally, encompassing human, animal, plant, and environmental health priorities, science, and communities.

Introduction

The Threat of Emerging Infectious Disease Outbreaks

Infectious disease outbreaks have surprised humanity throughout history. New diseases appear without warning, even today. Outbreaks of familiar diseases regularly occur where and when they are not expected.

During 2015-2016, an outbreak caused by the Zika virus has swept across much of the Western Hemisphere, its transmission driven primarily by the widely-distributed *Aedes aegypti* mosquito. Zika virus infection during pregnancy has resulted in thousands of devastating birth defects, and the fear of many more to come. In August 2016, the United States Government declared a public health emergency in Puerto Rico, which has experienced more than 32,000 Zika cases as of December 2016, including more than 2,000 pregnant women. Transmission is occurring also in the U.S. Virgin Islands, and limited local transmission occurred in Florida and Texas. Researchers discovered the Zika virus almost 70 years ago in Africa, but it received little attention before the current outbreak. Until recently, Zika was an obscure pathogen that caused mild disease in endemic settings or in occasional outbreaks that died out.

The Zika outbreak is just the latest public health emergency caused by an infectious disease. West Africa will long be recovering from the Ebola outbreak of 2014-2016, which killed approximately 11,000 people in the region and reached the United States and Europe. Although Ebola virus has also been known for decades, prior outbreaks had never been known to affect more than 500 people. Middle East Respiratory Syndrome Coronavirus (MERS-CoV), first identified in 2012, spread to dozens of countries worldwide, infected approximately 1,800 people, and caused death in one-third of those infected. A novel and virulent influenza strain could arise at any time and cause a catastrophic pandemic. The most severe, in 1918, caused an estimated 50-100 million deaths; a severe influenza pandemic today could possibly kill at least as many (despite medical and public health advances) (Osterholm, 2005) and bring about a global recession costing more than $3 trillion (World Bank, 2008).

The number, per year, of human infectious disease outbreaks and of diseases causing outbreaks has increased since 1980, according to one study (Smith et al., 2014). These recent outbreaks involve novel emerging diseases like MERS and Severe Acute Respiratory Syndrome (SARS), as well as known diseases that are re-emerging—appearing in new areas, developing resistance to treatments, or otherwise behaving in unfamiliar ways.

Contributing to the emergence (or re-emergence) of infectious disease of humans is a convergence of factors, including ecosystem alteration, urbanization, agricultural intensification, climate change, and globalized travel and trade (Morse, 1995; Smolinski et al., 2003). Especially important are factors that promote connections across humans, animals, and the environment (or ecosystem), enabling microbial spillover from natural reservoirs to humans. Zika, Ebola, MERS-CoV, and pandemic influenza viruses are human pathogens that originate in animals (“zoonotic” pathogens), as are most emerging and re-emerging infectious diseases of humans (Woolhouse and Gowtage-Sequeria, 2005; Appendix 1). Since 2003, the United States has used Emergency Supplemental appropriations 6 times to support Federal response to zoonotic viruses: for SARS-CoV (2003); avian or pandemic influenza (2005, 2006, and 2009), Ebola (2015), and Zika (2016).

Some of the factors contributing to human disease outbreaks also help drive infectious disease outbreaks of plants and animals, which are a threat to health, food security, economies, and ecosystems. For example, in 2014-2015, viral reassortment events combining North American and Eurasian avian influenza strains led to the emergence of novel strains that spilled over into commercial and backyard poultry operations across the United States. It was the largest animal health disaster in U.S. history, with removal and disposal of nearly 50 million birds. Approximately $879 million was spent on response activities and preparing for reemergence of the disease (Johnson et al., 2016).

Public health response to emerging infectious disease threats has often been largely reactive—a response is mounted after an outbreak is recognized. The United States need not, and should not, accept this *status quo*. Recent advances in science and technology (S&T) provide opportunities to mitigate large-scale outbreaks by predicting more accurately when and where outbreaks are likely to occur, and how they will progress. These S&T advances will not provide certainty, and will be constrained by limitations in scientific knowledge and epidemiological data, especially for novel diseases. But, if translated effectively into real-world applications, they can support better decisions about how, when, and where to apply public health resources to limit disease impact.

This report describes how the United States can harness S&T to advance the development and implementation of models to predict and mitigate the impacts of infectious disease outbreaks. The report provides an overview of the emerging field of outbreak modeling, surveys current Federal initiatives, describes challenges, and identifies possible Federal directions to surmount the challenges.

Infectious Disease Outbreak Modeling

Public health scientists rely increasingly on computational tools to understand epidemiological patterns and guide prevention and control measures (Heesterbeek et al., 2015). A new, interdisciplinary science is forming around the development of these approaches, spurred by scientific advances across a range of areas (including pathogen biology, genomics, bioinformatics, ecology, and machine learning), as well as near-real-time availability of relevant data during outbreaks (including traditional epidemiological data as well as newer data types such as pathogen genetic sequences, human mobility patterns, and social media data).

One especially promising area of this new field of computational biosciences is epidemiological predictive modeling—computer-based mathematical approaches to project the course of an outbreak that is underway, or to anticipate disease emergence or outbreaks (Box 1).

Box 1. Overview of approaches to outbreak modeling.

These models use diverse methods and data sources. One way of classifying them (simplistic though often helpful) is to view them as phenomenological or as mechanistic models. Phenomenological models describe or predict outbreak dynamics with no or little use of explicit assumptions about disease transmission. Examples include extrapolation of an epidemic curve using simple mathematical equations, and complex machine learning methods. Mechanistic approaches, in contrast, model the process of disease transmission. These also range from simple to quite complex, with dozens of parameters relating the state of the model at a given time to its future state, which must be estimated from the data or imputed. Traditional mechanistic models group classes of individuals together; computational advances now enable individual-based models for large populations, which model individual-level behavior and interactions.

Many types of descriptive and predictive models were used during recent, large-scale outbreaks. They addressed one or more of the following five uncertainties, which are the typical applications of modeling for outbreaks:

1. Outbreak growth rate and future course. A key parameter describing outbreak dynamics is the reproduction number, or the number of entities (individuals or sub-populations, such as herds or flocks), on average, infected by each infected entity. Early-outbreak estimates of the reproduction number provide an initial assessment of how rapidly the outbreak is growing; repeated over time, estimation of the reproduction number can help gauge the effectiveness of the response. A related modeling application is near-term forecasting of important public health quantities, such as the number of infections, clinical cases, or deaths.
2. Geographic spread. Datasets on travel and trade allow modelers to assess pathogen importation risk (e.g., for airports, farms, or ecosystems), identify areas at high risk for importation, and evaluate the utility of screening, or travel or trade restrictions. Such models may be coupled with disease transmission models to assess the chance that pathogen importation could spark a local outbreak. Other types of mobility data, such as anonymized mobile phone records, also have facilitated important insights into the role of population flows in disease outbreaks. More granular, empirical insights into disease transmission networks are possible with genomic pathogen surveillance during outbreaks; when coupled with standard epidemiological data, it can identify transmission links across spatially-distributed areas.
3. Possible interventions and clinical trials. A major contribution of outbreak modeling is assessment and comparison of possible control measures, such as limiting large gatherings, tracing contacts of patients to identify infections early, distributing drugs or vaccines, and culling animals on infected farms. Modeling allows such interventions to be evaluated in simulated environments before resources are devoted to implementing them. Models also allow researchers to determine the feasibility of launching large-scale field trials to evaluate new vaccines and therapeutics while the outbreak is underway.
4. Pathogen evolution and genetic determinants of function. As genomic data becomes available on pathogens circulating during an outbreak, models can establish how the outbreak strain is related to any strains that had previously been catalogued. These models can help in determining when the outbreak began, provide clues to important pathogen characteristics, and inform public health decisions that reduce the impact of

the outbreak. Models can use genomic surveillance data to detect evolutionary changes during the outbreak that could indicate further adaptation to new host species (animal or human) or vectors, which could portend more extensive transmission.

5. Pathogen origin and drivers of emergence. Few models attempt to predict the time and place of an outbreak beforehand, but retrospective studies of outbreak precursors have identified possible precipitating factors, and a broad class of models attempts to identify settings that could be conducive to disease emergence or spread (Box 2). Operationally useful models have been developed for some climate-sensitive diseases, such as those transmitted by arthropod vectors whose abundance and behavior are tightly linked with temperature, precipitation, vegetation patterns, and other aspects of weather or the environment. For zoonotic diseases, studies in areas of disease emergence that identify wildlife or domestic animal species, the microbes they carry, and animal or human behaviors that could promote spillover have been especially revealing.

Box 2. Predicting disease risk: Example of plant pest phenology and climate suitability.

These models are based on the biology of the specific pest and match the pest biology with climate and host characteristics to support plant pest detection activities. U.S. applications include models for various pests, such as *Dendrolimus pini* (pine tree lappet moth), *Epiphyas postvittana* (light brown apple moth), *Helicoverpa armigera* (Old world bollworm), *Monachamus alternatus* (Japanese pine sawyer beetle), *Platypus quercivorus* (oak ambrosia beetle), *Thaumatotibia leucotreta* (false codling moth). These pests can attack numerous host species, some (such as the old world bollworm) with hundreds of plant host species, including row crops (such as wheat, corn and cotton), vegetables (like peppers and tomatoes), and fruit crops (strawberries and stone fruits).

Modelers drew on diverse data sources to generate insights into these uncertainties during the recent Ebola outbreak in West Africa, the current Zika outbreak in the Americas, and non-human disease outbreaks in the United States, such as highly pathogenic avian influenza, white-nose syndrome in bats, and sudden oak death. Extensive modeling occurred during some of these emergencies (Box 3).

Box 3. Modeling to support recent outbreak responses.

Ebola: When the West Africa Ebola outbreak gained international attention in March 2014, initial reports warned that it was already spiraling out of control, but data collection and reporting challenges hampered assessment of the situation. Key knowledge gaps included the importance of various transmission settings and the virulence of the outbreak virus strain. Preliminary models incorporated the limited available data to estimate the growth rate of the epidemic and predict its course in the absence of public health intervention. These early forecasts were cited widely in calls for a major international response. Analyses of viral genomic data and epidemiological data suggested that the outbreak had begun with the introduction of a new Ebola virus strain in late 2013, and showed that human travel was propagating the outbreak across urban centers and national borders. As the outbreak continued, detailed mechanistic models tested various possible interventions, and suggested that control could most quickly be achieved with a combination of hospital- and community-based measures, such as safe burial practices and prohibition of eating and serving bushmeat. Models also were used to support planning of anticipated field trials of candidate vaccines against Ebola.

Zika: The Zika outbreak in the Americas also spurred a vigorous response by the modeling community, like the 2014-2016 Ebola outbreak did. As with Ebola, models were used to predict disease burden and international spread and define the genetic lineage of the outbreak virus strain, and contended with knowledge limitations similar to those encountered in the Ebola response. Models also were used to address additional uncertainties around the Zika epidemic. For example, they assessed the risk of international spread associated with the 2016 Summer Olympics in Rio de Janeiro, Brazil; the risk of local transmission by competent mosquito vectors if infected travelers introduce the virus into new areas; the resources needed to screen donated blood for Zika virus; and the risk of birth defects in infants born to women infected during pregnancy.

Highly Pathogenic Avian Influenza: In support of the response to the 2014-2015 outbreak of highly pathogenic avian influenza in the United States, models were developed to simulate disease introduction and spread within U.S. avian migratory flyways. These models were used to evaluate alternative control strategies for their impact on outbreak severity, duration, and economic consequences, including trade losses. This enabled improvements in surveillance design, planning for business continuity, evaluation of vaccination strategies, resource planning for diagnostic testing, and stockpiling of personal protective equipment and antiviral drugs. Phylogenetic models were used to better understand disease spread and inform efforts to improve biosecurity on farms.

Federal Efforts in Outbreak Modeling

The White House and Federal agencies have recognized the promise of modeling to improve epidemic preparedness and response. The policy foundation for Federal outbreak modeling includes the *National Strategy for Pandemic Influenza Implementation Plan* (2006), which identified models as “powerful tools that can be used to inform policy decisions by highlighting the impact of various interventions on the spread of disease”; *National Strategy for Biosurveillance* (2012), which included forecasting as a core function of biosurveillance; and *National Biosurveillance Science and Technology Roadmap* (2013), which noted that “New modeling and ecological forecasting approaches have the potential to enhance the effectiveness of current strategies for predicting the likelihood of disease outbreaks and determining the likely impacts when a threat is detected.”

To advance national capabilities for outbreak modeling, the National Science and Technology Council in 2014 chartered an interagency group, the Pandemic Prediction and Forecasting Science and Technology Working Group (PPFST WG). The PPFST WG provides Federal agencies a forum to coordinate priorities and activities in outbreak modeling. Many of its participating agencies support programs in the area, ranging from foundational research into the ecological and biological factors driving disease emergence to operational modeling that supports decisions during outbreaks (Table 1).

Table 1. Selected examples of recent Federal programs for infectious disease modeling.

Department or agency	Examples
Department of Agriculture	Animal and Plant Health Inspection Service: Foreign animal disease (e.g., avian influenza, foot-and-mouth disease) transmission modeling to guide preparedness and outbreak response. Plant health modeling to identify potential pathogen niches and to assess spatio-temporal pest introduction risk (e.g., the Spatial Analytics Framework for Advanced Risk Information Systems [SAFARIS] allows for modeling of plant pests and diseases globally and under climate change; the SAFARIS Census

	Travel Model tracks international passenger movement and identifies potential areas for plant pest and disease introduction). Agricultural Research Service: Research and modeling of arthropod vectors of human and animal disease and insect pests of crops.
Department of Defense	Defense Advanced Research Projects Agency: Modeling zoonotic spillover and predicting transmission of human diseases of military and national security importance. Defense Health Agency: *Global Emerging Infections Surveillance* (emerging infectious disease surveillance and outbreak response, including phylogenomic analysis of viral spread and forecasting of vector-borne, water-borne, and respiratory diseases). Defense Threat Reduction Agency: *Biosurveillance Ecosystem* (Chemical and Biological Technologies Department; system and predictive analytics development to enable real-time biosurveillance), *National Countering Weapons of Mass Destruction Technical Reachback Enterprise* (technical reachback support for weapons of mass destruction threats, including biological threats).
Department of Energy/National Laboratories	*Biosurveillance Ecosystem* (funded by Defense Threat Reduction Agency); intramural efforts related to infectious disease modeling.
Department of Health and Human Services	Biomedical Advanced Research and Development Authority: Intramural real-time epidemic analysis and forecasting of resource needs for influenza, Ebola, Zika; *Modeling Coordination Group* (forum for sharing of modeling results among interagency and academic communities). Centers for Disease Control and Prevention (CDC)*: Epidemic Prediction Initiative* (community to connect research on epidemic prediction to public health decision-making), *Health Economics and Modeling Unit* (develops and releases innovative, practical tools to support modeling needs within CDC and for the broader public health community). National Institutes of Health: *Models of Infectious Disease Agent Study* (collaboration of research and informatics groups to develop computational models of the interactions between infectious agents and their hosts, disease spread, prediction systems and response strategies), *Research and Policy in Infectious Disease Dynamics* (funding by Department of Homeland Security; workshops, working groups, fellowships to advance mathematical modeling of infectious diseases).
Department of Homeland Security	S&T Directorate: *Research and Policy for Infectious Disease Dynamics* (workshops, working groups, fellowships to advance mathematical modeling of infectious diseases).
Department of the Interior	U.S. Geological Survey*:* Intramural expertise across eight science centers in modeling of infectious diseases in aquatic and terrestrial wildlife, with an emphasis on understanding impacts on natural resources; includes the *National Wildlife Health Center* (bureau lead for terrestrial wildlife disease surveillance, which maintains high-biocontainment laboratories, is a World Organisation for Animal Health collaborating center, and

	conducts research and transmission modeling for diseases in wildlife, including avian influenza, plague, and other zoonoses).
Environmental Protection Agency	Office of Research and Development: Risk modeling of infectious diseases (vector-borne and biothreat agent) using environmental and socio-economic predictors.
Intelligence Advanced Research Projects Activity	*Mercury* (forecasting disease outbreaks using large volumes of data streams and ensemble techniques).
National Aeronautics and Space Administration	Earth Science/Public Health Application Area: Application of earth observations to public health, particularly for infectious diseases and other priority areas.
National Science Foundation	Directorate for Biological Sciences: *Ecology and Evolution of Infectious Diseases* (research on the ecological, evolutionary, and socio-ecological principles and processes that influence transmission dynamics of infectious diseases).
U.S. Agency for International Development	*Emerging Pandemic Threats* (regional, national, and local One Health capacity-building, including detection and discovery of zoonotic diseases at the wildlife-human interface).

Several of these agencies re-direct or focus efforts when new infectious disease threats arise, such as during the current Zika epidemic (Box 4).

Box 4. Selected examples of the Federal outbreak modeling response to Zika.

Operational modeling and data- and information-sharing. Biomedical Advanced Research and Development Authority performed real-time epidemic analysis to estimate demand for diagnostics and treatment resources, and convened the Modeling Coordination Group with Federal and non-Federal partners regularly to promote sharing of data and modeling results. Centers for Disease Control and Prevention estimated the risk for severe outcomes, such as microcephaly, and established a public website for sharing of machine-readable epidemiological data.

Modeling-related research & development. National Institutes of Health supported studies that modeled possible Zika control strategies; co-risk and interactions of Zika and other arthropod-borne viruses; genetic determinants of Zika emergence and fitness; competence of *Aedes aegypti* and *Aedes albopictus* to transmit Zika; and possible mosquito vector insecticide resistance, among other research. National Science Foundation funded rapid-response grants to develop predictive models and discover principles governing transmission dynamics for Zika. U.S. Geological Survey provided technical assistance for surveillance of the non-native *Aedes* vectors on Department of the Interior lands; supported studies to investigate the role of wildlife in Zika dynamics in the Americas; and developed flavivirus transmission risk models using climate, remote sensing and other geospatial data.

Challenges and Recommendations

Agencies across the Federal government are devoting research and development (R&D) funds to outbreak modeling, reflecting the broad importance of these capabilities across diverse missions. Potential synergies abound across this Federal landscape, and between Federal and non-Federal partners. At the same time, the vigorous application of modeling to recent outbreaks, within and beyond the Federal government, demonstrates a hope across varied scientific and public health communities that these models can help mitigate the impact of outbreaks. The field is poised to transform public health preparedness and response. But it faces challenges in achieving the vision of epidemic prediction. The following sections describe the three highest-priority challenges, and recommends Federal steps to address them.

Data- and information-sharing. Obtaining timely and accurate data and information during outbreaks has long been a major challenge to effective response. Today, though, the demand for timely and accurate data and information during outbreaks appears greater than ever. Outbreaks can quickly spiral out of control in a highly interconnected world; decision-makers need high-quality data and information, in near-real-time, to keep pace. This need is especially important at the beginning of a novel disease outbreak, when knowledge about the pathogen and data on the epidemiological situation is limited.

The outbreak modeling community therefore expends considerable effort in obtaining, cleaning, and updating epidemiological data that are not otherwise available in a readily usable form. During the Ebola and Zika outbreaks, for example, scientists transformed official Portable Document Format (pdf) reports from health ministries into machine-readable datasets, and shared them online (Rivers, 2015; CDC, 2016a). These datasets were used widely in modeling applications, and enabled studies by groups that could not have devoted the resources required to generate those more-usable data files.

The need for timely data extends beyond the epidemiological data traditionally collected during outbreak responses. Innovative models incorporate data on diverse potential predictors, such as spatio-temporal data on population locations, movement of hosts of disease agents or locations of vectors, environmental conditions, human behavior, and public health interventions. Like officially-reported outbreak surveillance data, such data may be available online, but may be scattered around multiple platforms and in different formats.

Another, fundamental barrier to outbreak modeling is failure to share data and information. Various public health and research groups collected epidemiological data as part of the responses to Zika, Ebola, and other major outbreaks. Some withheld data until they could publish results, even though the data may have been useful earlier for others to assess potential response efforts.

Some who did not share data may not have been aware of how others could use the data to support the response, or may have felt constrained by privacy considerations or by terms of agreement with data providers. For example, some Centers for Disease Control and Prevention operational modeling projects during the Ebola response were delayed because the required data-sharing agreements were not in place (Meltzer et al., 2016). Security and intellectual property concerns also may inhibit data sharing during outbreaks.

A community of interest is building around the need for wider and more rapid sharing of epidemiological data and information during public health emergencies (Wellcome Trust, 2016; WHO, 2015a). But the sharing failures of recent outbreaks show that much more needs to be done.

A PPFST WG survey of WG members and modelers participating in the National Institutes of Health Models of Infectious Disease Agent Study also provides evidence that incomplete access to data and information impedes the development and application of outbreak models (Appendix 2). Among respondents, 56% indicated that lack of access to existing data often or always delays their work; 33% reported that lack of access often prevents completion of a modeling project; and 44% had delayed or halted outbreak prediction or response efforts because existing data could not be obtained. Respondents indicated that if existing data were more readily available and usable, researchers could create more accurate, better resolved spatio-temporal models of infectious diseases to predict phenomena such as Ebola hotspots. High-priority data sources included data from electronic health records and pathogen genomic sequence data. One third cited technical considerations such as data formatting, standards, procedures, or means of dissemination as a hindrance to their analytical efforts.

To enable outbreak modeling that provides decision support based on relevant, timely, and accurate data and information, the Federal government should extend and adapt broader efforts around data science, open data, and open science (e.g., Executive Order 13642, Making Open and Machine Readable the New Default for Government Information, 9 May 2013) to data pertaining to infectious disease outbreaks. Specifically, the WG recommends that the Federal government, to the extent possible within existing authorities and resources:

Recommendation 1: Identify key questions likely to arise during outbreak response decision-making to help define and prioritize data collection and modeling goals.

Recommendation 2: Adopt a common plan for identifying the minimal essential, publicly-releasable data to support outbreak decision-making; and for rapidly processing, updating, and sharing these data publicly in a form readily usable for analysis by computational tools.

Recommendation 3: Collaborate with professional organizations, state and local public health offices, and other stakeholders to develop standards for collecting and sharing the types of data needed for modeling and analysis during outbreaks.

Recommendation 4: Ensure that data obtained with Federal funds that may address key questions during outbreak response decision-making is shared rapidly with other Federal stakeholders and in a form readily usable for analysis with computational tools (and with meta-data that describes potential restrictions).

Recommendation 5: Ensure that analytical products (e.g., scientific manuscripts) generated with Federal funds that may address key questions during outbreak response decision-making, or inform remediation efforts following such an emergency, and that are appropriate for public release, are shared rapidly with the public free of charge.

Recommendation 6: Develop S&T solutions to the privacy, security, and related challenges that impede data access, such as de-identification algorithms and synthetic datasets.

Outbreak model development and decision support. The extent of modeling during the current Zika outbreak in the Americas and recent Ebola outbreak in West Africa is noteworthy. More than 60 Ebola modeling analyses were published during the first 20 months of the outbreak (Chretien et al., 2015). An interim literature review by the PPFST WG identified 36 Zika modeling papers (published or as pre-prints) in the 7 months after the World Health Organization declared a Public Health Emergency of International Concern. This abundance of predictions and insights creates challenges for those making decisions about the response.

Merely keeping track of new modeling analyses during an outbreak is a demanding task. And, as best as the PPFST WG can determine, it is a mission that no organization has, within or outside of the Federal government. There seems to be no enduring, systematic effort anywhere to identify new modeling studies as they become available; evaluate methodologies; assess the validity of the approaches; and systematically synthesize results during outbreaks to update estimates of key epidemiological parameters and predictions, such as the reproduction number, the effectiveness of interventions, areas that may become affected, and the future course of the outbreak. Because all models rely on assumptions, efforts such as these to synthesize results across multiple models that are built on different assumptions help ensure robust conclusions.

A retrospective study of forecasts issued during the West Africa Ebola outbreak illustrates the need for such assessments: predictions of future incidence made at about the same time with about the same prediction time horizon differed 5- to 10-fold early in the outbreak (Chretien et al., 2015). As these varied predictions were issued during the outbreak, many public health analysts and decision-makers expressed uncertainty about which ones were more credible.

Cross-model comparisons are more feasible if the code underlying models used during outbreaks is released, so models can by implemented using standardized data and assumptions. But sharing of code for outbreak models may be even less common than sharing of other research products or results. For example, the interim PPFST WG assessment of modeling for the Zika outbreak found that all studies were made freely available online at the time of publication, but only one-fifth included a link to source code with the publication.

An example of a promising tool that could help facilitate model comparisons is the Biosurveillance Analytics Resource Directory (BARD), developed by Los Alamos National Laboratory and funded by the Defense Threat Reduction Agency (Margevicius et al., 2016). The BARD, however, does not house the models themselves and, to the PPFST WG's knowledge, no organization is charged with updating it or some similar repository as modeling accelerates during outbreaks.

Even with a systematic effort to track modeling results during outbreaks, discerning the more valid approaches, and thereby identifying the predictions and inferences warranting higher confidence, would be challenging. The outbreak modeling community has not adopted standards for assessing and reporting model attributes or performance. For some pathogens, human data that would be useful in building or validating models are limited or non-existent. And, in a dynamic scientific discipline, with new methodologies and data sources frequently introduced, the best approach in a given epidemiological scenario often is uncertain.

Controlled, multi-center modeling contests and projects, such as those organized recently by Federal agencies (Box 5) are generating valuable insights (e.g., simple models often perform as well as more complex models; model ensembles often perform better than any individual model), though there have been too few efforts like these to identify the best approaches across a range of likely scenarios.

Box 5. Recent federal outbreak prediction competitions.

Centers for Disease Control and Prevention (CDC): Seasonal Influenza Forecasting. The CDC Influenza Division has worked with external researchers since 2013 to improve the science and usability of influenza forecasts by coordinating prediction challenges for seasonal influenza in the United States (Biggerstaff et al., 2016). CDC's FluSight website (https://predict.phiresearchlab.org/), launched in 2016, displays the weekly forecast of each participating group, allowing comparison of the predictions to actual disease activity during the current influenza season. This effort is ongoing and is one of the few operational forecasting systems for an infectious disease in the United States.

CDC, Department of Defense (DoD), National Oceanic and Atmospheric Association (NOAA): Dengue Forecasting Project. Several agencies and the PPFST WG coordinated a public contest in 2015 to predict dengue incidence in Puerto Rico and Peru. CDC and DoD released historical dengue surveillance data that had not previously been made public for participants to use in model training. NOAA provided climatological data for consideration in predictive models. The PPFST WG held a workshop at the White House with teams that submitted the best-performing models to identify lessons for outbreak predictive modeling.

Defense Advanced Research Projects Agency (DARPA): Chikungunya Virus (CHIKV) Challenge. DARPA launched the CHIKV Challenge in 2014 as the virus swept through the Americas. Teams from around the world competed to develop the most accurate predictions of CHIKV cases for all Western Hemisphere countries and territories. Winners were selected based on overall accuracy and in other domains, including the tool's applicability and presentation. Los Alamos National Laboratory performed independent analysis of the entries.

National Institutes of Health (Fogarty International Center): Ebola Modeling Exercise. Teams from U.S., U.K., and Canadian universities and several federal agencies predicted the progress of the 2014-2016 West Africa Ebola epidemic under various containment scenarios; results were compared in a collaborative way to identify general features of more accurate methods and optimal approaches for different epidemiological settings.

Department of Agriculture: Model Validation Exercise. Teams from the United States, Australia, Canada, and New Zealand developed model parameter sets and scenarios to compare the predictions made by three simulation models of foot-and-mouth disease in use by these countries, and the implications of modeling results for decision-making. This approach of relative validation was identified as a way to increase end-user confidence and improve understanding of how model assumptions influence outcomes and recommendations (Dube et al., 2007).

Another promising research area for model development and assessment is the use of intensive, active surveillance programs, such as Columbia University's Virome of Manhattan project (funded by the Defense Advanced Research Projects Agency and National Institutes of Health; Shaman, 2016). These testbeds may provide important insights not obtainable from routine public health surveillance. Incorporation of genomic sequence data; human, animal, and environmental assessments; and machine learning and other promising approaches in predictive model development could make such efforts especially powerful.

To support development of outbreak models, and identification and synthesis of results across the most credible models, the Federal government should advance scientific and operational capacities for meta-analysis of outbreak modeling. Specifically, the WG recommends that the Federal government, to the extent possible within existing authorities and resources:

Recommendation 7: Enable Federal efforts to systematically identify outbreak modeling studies as they become available; conduct controlled comparisons of models if source code is accessible; assess the quality of models and modeling studies; and maintain current estimates of key epidemiological and pathogen parameters through synthesis of modeling results.

Recommendation 8: Strengthen communication and coordination between modeling groups and Federal offices requiring outbreak decision support.

Recommendation 9: Strengthen support for development of standards or best practices to assess outbreak model attributes and performance.

Recommendation 10: Strengthen support for controlled comparisons of outbreak modeling approaches to identify preferred approaches for possible real-world scenarios.

Recommendation 11: Strengthen support for development of decision-support methods and tools to facilitate synthesis and adjudication of results across multiple models.

Recommendation 12: Establish criteria for public release of outbreak model code developed with Federal funds, and ensure that code that is appropriate for public release is shared publicly, free of charge, no later than the time that the scientific paper describing the model is published.

Recommendation 13: Establish active, population-based surveillance testbeds for rigorous development, comparison, and validation of infectious disease models (with first consideration given to building modeling components into existing epidemiological surveillance or research efforts).

<u>Science of disease emergence.</u> The two challenges described above pertain mainly to modeling conducted after an outbreak has begun (or to a hypothetical outbreak). The third—understanding the processes that drive disease emergence and transmission well enough to predict where and when diseases are likely to emerge—may be an even greater challenge. Achieving this capability, however, promises tremendous rewards in life, economic prosperity, and security.

Most emerging infectious diseases affecting humans, including Zika, Ebola, MERS, pandemic influenza, SARS, and other recent outbreaks of regional and global importance, are zoonotic. Often, novel zoonotic pathogens are discovered only after they have caused outbreaks among humans, when searches for the particular pathogen are targeted to candidate animal reservoirs. The zoonotic pathogens that have spilled over into humans, though, represent only a small fraction of the pool of potential pathogens circulating in animals. The size of this pool is unknown, but one estimate by Federal scientists is that there are around 500,000 yet-to-be-discovered viruses capable of infecting humans (Carroll, 2016). As the boundaries between humans and animals decrease, the risk of zoonotic disease emergence will continue to increase.

Integrated, human-animal modeling studies and efforts to discover potential pathogens are critical for understanding the process of pathogen spillover from animals to humans. An example is the U.S. Agency for International Development Emerging Pandemic Threats (EPT) program, initiated in 2009 to pre-empt zoonotic disease outbreaks or mitigate their impacts. In collaboration with the Centers for Disease Control and Prevention, the Smithsonian Institution, and non-Federal partners, EPT supports projects in 20 countries to build regional, national, and local "One Health" capacities for early detection, laboratory-based diagnosis, rapid response and containment, and risk reduction. One component of EPT, PREDICT, includes discovery of zoonotic diseases at the wildlife-human interface and development of predictive models for disease emergence and spread. PREDICT has sampled more than 56,000 non-human primates in countries around the world and discovered more than 815 novel viruses.

Discovery of potential zoonotic pathogens is not sufficient, however, to understand the risk they may pose to humans. There have been too few spatio-temporal surveys to understand transmission dynamics, particularly across multiple interacting species (for example, wildlife, nearby domestic animals, and humans); and investigations of transmission dynamics between humans and animals to which exposure may be especially high, such as companion animals and animals used for food. Little also is known about how to predict whether a novel agent in a non-human reservoir possesses (or could acquire) the capability to infect humans, cause disease, and be transmitted among humans; or about the interplay of biological, ecological, environmental, and social-behavioral factors in driving disease emergence.

To advance public health towards the vision of pandemic prediction, the Federal government should strengthen support for modeling R&D, especially in animals and at the human-animal-environment interfaces. Specifically, the WG recommends that the Federal government, to the extent possible within existing authorities and resources:

Recommendation 14: Develop a "One Health" national strategy for predictive modeling of potential infectious disease threats globally, encompassing human, animal, plant, and environmental health priorities, science, and communities.

Urgency of Action

The next pandemic pathogen could emerge any day. Perhaps it already is circulating in a non-human reservoir, but has not yet had the opportunity to infect humans. Changes in climate, ecosystems, or human settlements and behavior could provide that opportunity. Or perhaps it will arise from a precursor microbe, after poorly-understood evolutionary drivers bring about the genomic changes necessary to infect and be transmitted among humans. The next outbreak to devastate agriculture or ecosystems also could emerge at any time, and spread quickly through a highly networked world. Interdisciplinary collaborations of mathematicians, computer scientists, microbiologists, epidemiologists, veterinarians, and other scientists to understand and predict these processes are not esoteric, academic pursuits. These collaborations are vital investments that could provide tangible benefits to public health, economic vitality, and security.

The needed collaborations stretch well beyond the expertise and resources of any one part of the Federal government, or of the Federal government as a whole. The PPFST WG will continue its efforts to bring Federal agencies together, and join then with non-Federal partners and civil society, to advance outbreak modeling S&T, and catalyze its implementation against the rising threat of emerging infectious diseases.

Appendix 1

Recent Outbreaks Among Humans Caused by Zoonotic Viruses

1997-present: Avian influenza (H5N1). The H5N1 avian influenza strain was discovered in birds in China in 1996, and caused the first known human cases in Hong Kong in 1997. The 18 cases appeared to have been infected by birds in live poultry markets; the outbreak ended with the slaughter of all poultry in markets and farms in Hong Kong. Variants of the strain circulated in poultry and wild birds in Asia, and caused disease in humans again beginning in 2003. Since then, more than 850 human cases have been reported from 15 countries, and about half died (there have been no reported infections in birds or humans in the United States). Most human cases have been associated with contact with infected birds, although limited human-to-human transmission likely has occurred rarely. In 2004-2005, human case clusters and widespread avian outbreaks in Asia prompted concerns of a pandemic and efforts globally to strengthen surveillance and response capacity (WHO, 2005).

2003: Severe Acute Respiratory Syndrome (SARS). A new disease first recognized in China in February 2003, SARS reached North America, South America, Europe, and other parts of Asia, with 8,096 cases and 774 deaths through July 2003. Most infections resulted from human-to-human transmission linked to hospitals; infection control measures ended the outbreak. SARS Coronavirus (SARS-CoV) originated from similar viruses in bats. It crossed the species barrier to infect civets and other animals in live-animal markets, which transmitted the virus to humans (de Wit et al., 2016).

2009-2010: Pandemic influenza (H1N1). Typically a seasonal disease, influenza type A viruses can cause pandemics when novel strains are transmitted from animals to humans and then circulate among humans. Wild aquatic birds are the natural hosts of influenza A viruses and the source of influenza viruses introduced into other mammals, such as pigs and chicken, which are, in turn, important sources of novel strains in humans. The most recent influenza pandemic, caused by an H1N1 strain in 2009-2010, is estimated to have caused over 250,000 deaths worldwide (Dawood et al., 2012) (for comparison, the most devastating influenza pandemic, in 1918, caused around 50 million deaths). The new 2009 strain resulted from reassortment of pig, bird, and human influenza strains.

2012-present: Middle East Respiratory Syndrome (MERS). A new disease first recognized in Saudi Arabia in 2012, MERS spread worldwide and continues to circulate as of September 2016, with around 1,800 cases, about one-third of them fatal. Most infections have resulted from human-to-human transmission linked to hospitals. Like SARS-CoV, MERS-CoV originated from similar viruses in bats. Its crossover to camels is estimated to have occurred more than 30 years ago. Unlike SARS-CoV, which infected but does not appear to have circulated among its intermediate non-human hosts (live-marked animals), MERS CoV transmission occurs among camels, resulting in more frequent introductions to humans (de Wit et al., 2016).

2013-present: Chikungunya. First identified in 1950s in East Africa, Chikungunya virus (CHIKV) is transmitted by *Aedes* species mosquitoes and infects humans as well as non-human primates and other small mammals. The disease is rarely fatal in humans, but may cause fever and debilitating joint pain. Chikugunya epidemiology changed recently, with outbreaks of unprecedented size in Indian Ocean islands and India, and the first local transmission in Europe and, in 2013, the Americas (with more than 1.7 million suspected cases in 45 countries, including local transmission in Florida, Puerto Rico, and the U.S. Virgin Islands, as of September 2016). CHIKV genetic adaptation to *Aedes albopictus* mosquitoes, which are widespread globally, may have contributed to chikugunya's expanding range (Coffey et al., 2014).

2014-2016: Ebola. Outbreaks of this viral hemorrhagic disease involve human-to-human transmission after infection by wild animals, such as apes; bats are suspected as the natural reservoir. The recent West Africa outbreak was the largest by far, with more than 28,000 cases and 11,000 deaths; limited transmission occurred in the United States and Europe after return of infected travelers. Previous outbreaks, dating to Ebola's discovery in 1976, occurred only in equatorial Africa and were much smaller (fewer than 500 cases each). Factors contributing to the West Africa outbreak include delayed recognition of a disease unfamiliar in the region, high population mobility across borders, weak public health infrastructure, and community resistance to control efforts (WHO, 2015b).

Zika: 2015-present. Discovered in East Africa in 1947, Zika virus (ZIKV) is transmitted primarily by mosquitoes (*Aedes*) and infects humans and other animals, primarily monkeys. Most patients have no or mild symptoms. Emerging evidence indicates a sexual transmission route, as well as rare neurological complications, including birth defects. Spillover from non-human hosts or limited transmission among humans caused small-scale outbreaks in Africa and Asia. The first large outbreak was reported in Yap Island (Federated States of Micronesia) in 2007, with more than 100 cases. Local Zika transmission was reported in Brazil in 2015, for the first time in the Western Hemisphere. The outbreak spread across the Americas and reached the United States, with locally-acquired cases in Puerto Rico (more than 13,000) and Florida (35) (CDC, 2016b). Factors that may explain the unprecedented size, spatial extent, and clinical impacts of the current outbreak remain under study.

Appendix 2

PPFST WG Survey: Access to Data for Outbreak Science

The Open Data action area, an interest group within the PPFST WG, has worked to characterize and address the effect of data availability on the ability of Federal agencies and collaborators to predict and respond to outbreaks. The Open Data group of PPFST developed and arranged for the administration of a data collection tool to learn more about the existing levels of access to data, effects of access on the ability to conduct outbreak science, and technical barriers to access. The survey was administered to members of PPFST as well as both Federal employees and academic researchers in the Models of Infectious Disease Agent Study of the National Institutes of Health during the week of 30 August 2016. Eighteen surveys were returned to the Open Data group of PPFST.

Survey respondents indicated that lack of access to data is severely hindering outbreak preparedness and response duties. A hindrance index score combining the responses to all survey items showed that lack of access to data created a hindrance to over half of outbreak science efforts, on average. The hindrance index scores were created on a scale from 0 (no hindrance) to 100 (maximum hindrance) and ranged from 18 (low hindrance) to 84 (high hindrance), with a mean of 53.

Lack of access to existing data leads to delays and in some cases complete obstructions of outbreak science efforts. Fifty-six percent of respondents indicated that data non-availability delays their outbreak science work “often” or “always”. Thirty-three percent of respondents said that data non-availability “often” entirely prevents them from completing outbreak science. Forty-four percent of respondents had delayed or deferred a decision on outbreak prediction and response efforts due to non-availability of data.

Respondents indicated low confidence in the accuracy of data, reporting that they are only confident in the accuracy of publicly available outbreak data less than half of the time on average. Respondents indicated that on average, the data they need to produce infectious disease outbreak models and analyses was publicly available in a timely fashion only less than half of the time. One practitioner noted that this lack of open data availability delays infectious disease short-term forecasting efforts.

Situational awareness and context information needed to properly interpret data inputs and build representative models are similarly lacking and were available to respondents through public channels less than half of the time.

Respondents indicated that if existing data were more readily available, then researchers could create more accurate, better resolved spatio-temporal models of infectious diseases to predict events such as Ebola hotspots. Outbreak genomic sequence information could help identify optimal diagnostic assays and treatment plans for newly-circulating strains. Access to BioWatch databases, electronic health records, and real-time medical claims data were suggested as critical components of outbreak science efforts.

One-third of respondents cited technical considerations such as data formatting, standards, procedures, or means of dissemination as a hindrance. Half of respondents suggested that the government should use a standardized format when releasing data such as outbreak reports publicly. In order to facilitate data distribution and use, about half of respondents suggested that the government should use a standardized format such as XML files for distribution, with a preference for no distribution restrictions.

Based on these findings, the Open Data group of PPFST recommends that further action be taken across the Interagency to expand public availability of detailed, highly spatially resolved population health and laboratory data in standardized formats in near real-time.

References

Biggerstaff M, Alper D, Drezde M, et al. 2016. Results from the Centers for Disease Control and Prevention's Predict the 2013-2014 Influenza Season Challenge. BMC Infect Dis 2016;16:357. http://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-016-1669-x

Carroll D. 2016. The Global Virome Project: the beginning of the end of the pandemic era. World Affairs. 13 Jul 2016. http://www.worldaffairs.org/blog/765-guest-post-the-global-virome-project-the-beginning-of-the-end-of-the-pandemic-era

Centers for Disease Control and Prevention (CDC). 2016a. Epidemic Prediction Initiative. GitHub. https://github.com/cdcepi

Centers for Disease Control and Prevention (CDC). 2016b. Zika virus. http://www.cdc.gov/zika/geo/united-states.html

Chretien JP, Riley S, George DB. 2015. Mathematical modeling of the West Africa Ebola epidemic. eLife 2015;4. https://elifesciences.org/content/4/e09186

Coffey LL, Failloux AB, Weaver SC. Chikungunya virus-vector interactions. 2014. Viruses 2014;6:4628. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4246241/

Dawood FS, Iuliano AD, Reed C, et al. 2012. Estimate global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modeling study. Lancet Infect Dis 2012;12:687.

De Wit E, van Doremalen N, Falzarano D, Munster VJ. 2016. SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol 2016;15:523.

Dubé C, Stevenson MA, Garner MG, Sanson RL, Corso BA, Harvey N, Griffin J, Wilesmith JW, Estrada C. 2007. A comparison of predictions made by three simulation models of foot-and-mouth disease. New Zealand Veterinary Journal 2007:55:6.

Heesterbeek H, Anderson RM, Andreasen V, et al. 2015. Modeling infectious disease dynamics in the complex landscape of global health. Science 2015;347:aaa4339. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4445966/

Johnson KK, Seeger RM, Marsh TL. 2016. Local Economies and Highly Pathogenic Avian Influenza. Choices. Quarter 2. http://www.choicesmagazine.org/choices-magazine/theme-articles/economic-consequences-of-highly-pathogenic-avian-influenza/local-economies-and-highly-pathogenic-avian-influenza

Margevicius KJ, Generous N, Abeyta E, et al. 2016. The Biosurveillance Analytics Resource Directory (BARD): facilitating the use of epidemiological models for infectious disease surveillance. PLoS One 2016;11:e0146600. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146600

Meltzer MI, Santibanez S, Fischer LS, et al. 2016. Modeling in real time during the Ebola response. MMWR Suppl 2016;65:85. http://www.cdc.gov/mmwr/volumes/65/su/su6503a12.htm

Morse SS. 1995. Factors in the emergence of infectious diseases. Emerg Infect Dis 1995;Jan-Mar:7.

Osterholm MT. 2005. Preparing for the next pandemic. N Engl J Med 2005;352:1839.

Rivers CM. 2015. Data for the 2014 Ebola outbreak in West Africa. GitHub. https://github.com/cmrivers/ebola

Shaman J. 2016. Virome of Manhattan. http://blogs.cuit.columbia.edu/jls106/projects/virome-of-new-york-city/

Smith KF, Goldberg M, Rosenthal S, et al. 2014. Global rise in human infectious disease outbreaks. J R Soc Inteface 2014;11:20140950. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4223919/

Smolinski MS, Hamburg MA, Lederberg J, eds. 2003. Committee on Emerging Microbial Threats to Health in the 21st Century. Board on Global Health. Institute of Medicine. Microbial threats to health: emergence, detection, and response. The National Academies Press. https://www.nap.edu/catalog/10636/microbial-threats-to-health-emergence-detection-and-response

Wellcome Trust. 2016. Sharing data during Zika and other global health emergencies. 10 Feb 2016. https://wellcome.ac.uk/news/sharing-data-during-zika-and-other-global-health-emergencies

Woolhouse ME, Gowtage-Sequeria. 2005. Host range and emerging and reemerging pathogens. Emerg Infect Dis 2005;11:1842. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3367654/

World Bank. 2008. Internal reported cited in: Center for Infectious Disease Research and Policy. World Bank says flu pandemic could cost $3 trillion. 17 Oct 2008. http://www.cidrap.umn.edu/news-perspective/2008/10/world-bank-says-flu-pandemic-could-cost-3-trillion

World Health Organization (WHO). 2005. Avian influenza: assessing the pandemic threat. Jan 2005. http://apps.who.int/iris/bitstream/10665/68985/1/WHO_CDS_2005.29.pdf

World Health Organization (WHO). 2015a. One year into the Ebola epidemic: a deadly, tenacious and unforgiving virus. Jan 2015. http://www.who.int/csr/disease/ebola/one-year-report/introduction/en/

World Health Organization (WHO). 2015b. Developing global norms for sharing data and results during public health emergencies. 2 Sep 2015. http://www.who.int/medicines/ebola-treatment/blueprint_phe_data-share-results/en/