CDC Foundation
Together our impact is greater

Modernizing Health Data Analytics and Forecasting

Forecasting and Modeling Listening Sessions

FINAL REPORT WITH TECHNICAL NOTES

Prepared by the Illinois Public Health Institute | August 27, 2021

Contents

CDC Foundation.
"Modernizing Health Data Analytics and Forecasting. Forecasting and Modeling Listening Sessions – Final Report with Technical Notes."
August 27, 2021.

Background and Report Overview

Real-time, actionable evidence serves as the cornerstone for improving baseline population health as well as preventing and responding to epidemics, outbreaks and pandemics. In January 2021, noting the importance of such evidence, the United States (U.S.) Government announced plans for establishing an interagency National Center for Epidemic Forecasting and Outbreak Analytics (hereafter "Center") to modernize global early warning and trigger systems to prevent, detect and respond to biological threats[1]

As an independent nonprofit that mobilizes philanthropic and private-sector resources to support the Centers for Disease Control and Prevention (CDC) and public health, CDC Foundation held two three-hour, multi-sector listening sessions to seek input on needs and gaps in public health forecasting. The listening sessions were held on July 22 and 28, 2021. These listening sessions engaged stakeholders across the private, public health, healthcare and academic sectors. Participants discussed ways to catalyze transformative advances in forecasting, mathematical modeling and other analytical capabilities in the United States and globally.

> *"You cannot predict the future, but you can create it."*
>
> —PETER DRUCKER

Discussions ranged across topics to assure broad feedback from the diverse subject matter experts in attendance, focusing notably on the nexus of forecasting, data access and data utility. Topics included 1) data collection, prioritization, integration and analyses; 2) short-term forecasts that may be unconditional predictions and longer-term forecasts that will be conditional predictions; 3) infectious and noninfectious diseases and syndromes; and 4) engagement and coordination across partners, sectors, stakeholders and jurisdictions.

This report summarizes high-level themes from the listening sessions and incorporates key takeaways from each theme. High-level themes included:

1. Modeling and forecasting
2. Infrastructure
3. Data collection and accessibility
4. Real-time and granular data
5. COVID-19 and other diseases
6. Surveillance
7. Partnerships
8. Data sharing
9. Communications

1 Biden, President Joseph. National Strategy for the COVID-19 Response and Pandemic Preparedness. Whitehouse.gov, 21 Jan. 2021, www.whitehouse.gov/wp-content/uploads/2021/01/National-Strategy-for-the-COVID-19-Response-and-Pandemic-Preparedness.pdf.

With that thematic approach, the report aims to provide multi-sector insights into important considerations for public health as leaders and practitioners work toward real-time, actionable evidence that can better meet ongoing and emerging needs of disease prevention and health promotion.

While the listening sessions yielded rich discussion resulting in the themes, this report represents the available input based on the current phase in the process. The nature of the data modernization work is dynamic. As additional opportunities for input are made available, new information will be added and recommendations may be refined.

Opening Remarks

At the start of both listening sessions, Judith Monroe, MD, President and CEO at CDC Foundation and Daniel B. Jernigan, MD, MPH, Deputy Director for Public Health Science and Surveillance at the CDC welcomed the group and provided background and level-setting remarks about the importance of advancing forecasting in the public health. The below summarizes those opening remarks.

JUDITH MONROE, MD, PRESIDENT AND CEO, CDC FOUNDATION

CDC Foundation is committed to identifying critical improvements needed to reinforce and support vital public health infrastructure for data modernization. Timely and actionable health-related data is the cornerstone for an effective response for the numerous conditions and diseases that negatively impact people's health and the economy. Forecasting, mathematical modeling and other analytical tools are valuable in interpreting and translating data into action to respond effectively to existing and emerging public health threats. In addition, the COVID-19 pandemic has demonstrated that this need for timely and actionable data is imperative for a broad range of decision makers, such as elected officials, institutional and organizational leaders, grassroots community organizers and others across a broad range of sectors, such as schools, workplaces, industry sectors and other institutions across communities. CDC Foundation sits at the intersection of the government, philanthropic and private sectors; working together these combined resources, insights and flexibility allow for tackling the toughest health challenges including modernizing health data.

DANIEL B. JERNIGAN, MD, MPH, DEPUTY DIRECTOR FOR PUBLIC HEALTH SCIENCE AND SURVEILLANCE, CDC

The Data Modernization Initiative falls under the purview of Public Health Science and Surveillance at the Centers for Disease Control and Prevention (CDC). Years of preparatory work is now being applied, with recent appropriations and funding intended to support public health data surveillance, analytics infrastructure and other modernization initiatives at CDC.

To date, there have been several challenges in addressing data modernization as part of the COVID-19 response, including:

- Limitations across mission critical public health responses such as detecting, tracking, intervening and preventing.
- Vastly different inputs leading to different estimates used to predict disease transmission and impact, forecast supplies and design countermeasures. There is not a coordinated effort and there are not enough benchmarks in place.
- Lack of coordination in the translation of models and forecasts to inform decision makers and to direct resources effectively. This has resulted in varying recommendations and efforts among jurisdictions.
- Limited and not easily accessible modeling and forecasting expertise.

The appropriations and funding directed at data modernization allows for establishing, expanding and maintaining efforts to modernize the United States' disease warning system to forecast and track hotspots for COVID–19 and its variants and other emerging biological threats, improving academic and workforce support for analytics and informatics infrastructure and improving data collection systems. To that end, the current Administration is looking to create an interagency National Center for Epidemic Forecasting and Outbreak Analytics. The goals for the Center include:

- Modeling and forecasting public health concerns and sharing information in real time to trigger governmental, private sector and public actions to respond within the United States and abroad.
- Advancing the use of forecast and outbreak analytics in public health decision-making with the goal of supporting more efficient and effective outbreak responses.
- Bringing together next-generation public health data scientists, expert disease modelers, public health emergency responders and high-quality communications experts to meet the needs of decision makers and translate information for key audiences.
- The proposed functions of the Center include:

Predict: Modeling and forecasting to determine the foundational data sources needed; support research and innovation in outbreak analytics and science for real-time action; and establish appropriate forecasting horizons. This includes surveillance and case data.

- Data inputs include public health departments, laboratories, health care, weather, animal outbreaks, social mobility and consumer data.

Connect: Broad capability for data sharing and integration; maximizing interoperability with data standards and utilizing open-source software and application programming interface (API) capabilities with existing and new data streams from the public health ecosystem and beyond.

Inform: Translating and communicating forecasts; connect with key decision makers across sectors including government, businesses and nonprofits, along with individuals with strong intergovernmental affairs and communication capacity for action. Take what can be forecasted and, through scenario modeling, translate that into action by decision makers and to inform the public.

High-level Cross-Cutting Themes

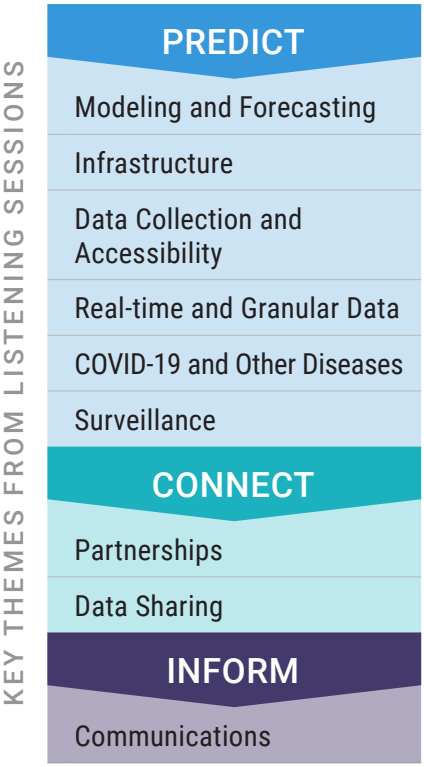

FIGURE 1

Themes that emerged across the listening sessions include:

- Modeling and forecasting
- Infrastructure
- Data collection and accessibility
- Real-time and granular data
- COVID-19 and other diseases
- Surveillance
- Partnerships
- Data sharing
- Communications

These themes (figure 1) are categorized within the three proposed functions of the Center described above.

The remainder of the report details the feedback received in the listening sessions on each of the themes. **Multiple participant comments related to the same idea are rolled up into one statement to support each theme.** Key takeaways were derived primarily from recommendations that are more specific and actionable and, in some cases, reflect participant consensus. For more information on the methodology of this report, see Appendix B.

PROPOSED FUNCTION

Predict

MODELING AND FORECASTING

Modeling and forecasting are analytical capabilities that can help guide policy and planning in responding to infectious diseases and other biological threats. Modeling and forecasting have the potential to improve epidemic and pandemic management by preventing deaths and severe illness and reducing the public health and economic impacts of these threats. Across the two listening sessions there were almost 200 comments related to this topic that aligned with the following subcategories: learnings from previous and existing models, precise and accurate interpretations, an integrated systems approach, individual vs. population-level forecasting and intentional matching between stakeholder needs and approaches to modeling and forecasting. Specific recommendations from each subcategory are highlighted below:

Learnings from previous and existing models

- Research, evaluate and learn from past and existing models, for example, weather forecasting models and applications, flu modeling and forecasting, MIDAS (Mixed Data Sampling) and RAPIDD (Research and Policy for Infectious Disease Dynamics). Also evaluate and learn from where things have gone wrong in generating forecasts. For example, flu forecasting could have brought different teams together from forecast hubs and other efforts.
- Evaluate approaches to the COVID-19 pandemic to assess the success of the various models and forecasts.
- Evaluate successful modeling and forecasting techniques implemented in other countries, such as Great Britain.
- Implement an impact assessment to gauge if particular models were/are effective.

Prioritize precise and accurate interpretations

- Communicate forecasting assumptions to stakeholders. Stochasticity that exists in the real world cannot always be emulated in a model. Similarly, it is important to present both the positives and limitations of the forecast. These can be taken into consideration for future forecasts.
- Create modeling systems that are more objective and less susceptible to bias.
- Recommend nowcasting over forecasting.

- Share whether forecast is counterfactual or factual in interpreting and communicating out the results.
- Interpret and use modeling and forecasting correctly and in a uniform manner.
- Create indicators/measures to assess accuracy of interpretations.

Integrated systems approach

- Develop an integrated systems approach and build out capacity and systems for simultaneous analytical practices, modeling and forecasting techniques.
- Coordinate among parallel modeling centers. There is currently extensive duplication in modeling and forecasting.
- Realign incentives with infectious disease modeling efforts in community and in academia.
- Look at what predictions and forecasts look like next to one another.
- Evaluate interagency links that will allow for more effective delivery of information that is wanted and/or important to local constituents.

The 2020 study *Ensemble Forecasts of Coronavirus DISEASE 2019 (COVID-19) in the U.S.* supports the above points by identifying that real-time ensemble forecasts, i.e. combining multiple probabilistic models and assessing forecast skills at different prediction horizons, can provide robust short-term predictions of relevant indicators to public health decision makers.[2]

Individual vs. population-level forecasting

- Ensure research capacity includes modeling at several levels, from ecological/population-level models to individual-based models. Ensure modeling incorporates variability in human actions and interactions that influence emergent social phenomena.
- Recommend individual-based models and network-based models that show how individual interactions influence the whole.
- Consider that local modeling requires local communication (local people translate data for local consumption).
- Recommend starting with forecasting goals that are meaningful and add value at the local level.

Intentional matching between needs and approaches

- Look at the key questions that need to be addressed and the best data that can support that modeling, prior to collecting data. Approaches to forecasting must be directly related to the question at hand.
- Clarify to all involved the prioritization, modeling targets and goals, and questions to be answered. Clarify what types of data are needed ahead of time.
- Have structured relationships that focus on how questions are addressed together.

2 Ray, Evan L, et al. "Ensemble Forecasts of Coronavirus DISEASE 2019 (COVID-19) in the U.S." *MedRxiv*, 2020, doi:10.1101/2020.08.19.20177493.

KEY TAKEAWAYS

In planning for the National Center for Epidemic Forecasting and Outbreak Analytics evaluate and emulate, as appropriate, existing forecasting models such as weather and flu modeling. The plan should include an integrated systems approach such that there is coordination across modeling centers and incentives are aligned. Ahead of collecting data, recommend intentional matching between the question at hand and approaches to and sources of collecting data. All relevant stakeholders should have clarity on modeling targets and goals. Consider modeling and forecasting at both individual and population-based levels. Assumptions and limitations must be communicated to those utilizing the forecasts.

INFRASTRUCTURE

"There is increasing consensus that infrastructure is crucial for connectivity, and that access to infrastructure is asymmetric. Therefore, tracking the natural buildup of physical infrastructure (and gaps in digital infrastructure) may serve as a crucial means for more rapid pandemic response."[3] Another top theme that emerged across the two listening sessions was the emphasis on infrastructure, with over 150 comments on this theme. Infrastructure in this context comprises resources, academic and workforce support and funding needed for the operations of the National Center for Epidemic Forecasting and Outbreak Analytics. Specific feedback on each of these areas included:

Resources

- Tap into existing resources and platforms until we can increase knowledge and understanding of modeling and forecasting.
- Prioritize what modernized data systems and infrastructures are needed for success, such as technology, a data lake, a system that links laboratory and electronic medical record (EMR) data, a data stockpile and/or a data hub.
- Recognize that smaller health departments may not have the capacity to work on urgent projects.
- Enhance the data pipelines in clinical infrastructure that enables faster reporting.
- Invest in global capacity development in modeling.

Academic and Workforce Support

- Grow the workforce across universities and other stakeholders. There is currently a workforce gap. Need capacity and expertise for different types of modeling, especially at the population level.
- Incorporate modeling and forecasting as topics into curricula for public health professionals and other professions.
- Have statisticians, epidemiologists, data analysts etc. observe and learn forecasting and modeling.
- Implement more research-based training programs and workgroups to support the workforce, including education around modeling.
- Consider incentives, particularly for academics and EMR custodians, for implementation.
- Offer workshops where modelers and frontline public health officials can come together to learn from one another.

3 Armanios, Daniel E, and Nicola Ritsch. *Infrastructure as Social Sensor: The Case for Better Collection and Integration of Infrastructure Data for Improving Pandemic Response.*

Funding

- Invest in operational structures, infectious disease analytics, modeling and forecasting. Data priorities are dependent on funding, investment and capacity. This work takes continued time, investment and communication.
- Build a structure during "peace" time such that it can be leveraged during concern periods.
- Fund data collection, aggregation and analysis on an ongoing basis.
- Invest substantially in state and local public health since they do not currently have the capacity to generate the needed data.

Operations

- Create operational structures to make technology and science work in a sensible way.
- Develop operational systems and integrated platforms in "peace" time for data curating, sharing, analytics, modeling and forecasting.
- Maintain all components in terms of people who are employed to do the day-to-day operational work, as well as the go-betweens (between the academics doing research and the operational side).
- Be aware that there is currently a lot of duplication in the work along with a lack of coordination and centralization of tools.
- Support interdisciplinary contributions in an ongoing way, with regular collaboration and feedback and an organized structure.
- Emphasize that multiple aspects of forecasting and modeling need to be a continuous, ongoing process.
- Ensure continuous and transparent dialogue, evaluation and feedback.
- Prioritize operations.

As noted in Armanios and Ritsch's report (see Appendix D), a focus on and enhancements to infrastructure can also be linked to improved economic and health outcomes and plays an important role in social determinants of health.

KEY TAKEAWAYS

Utilize existing tools and resources related to modeling and forecasting while growing the academic and workforce support. Offer workshops whereby modelers and frontline public health officials come together to learn from one another. Implement a data lake and a system that links laboratory and EMR data. Make short- and long-term investments in a more robust informatics workforce and maintain these investments, and the Center's capacity and readiness, even when there are no major events or crises. Prioritize operations.

DATA COLLECTION AND ACCESSIBILITY

As part of modeling and forecasting, participants gave special attention in their feedback to data collection at national and local levels, electronic medical records (EMRs) and healthcare data, social determinants of health data, geographic considerations, access to data (identified as the biggest bottleneck to progress) and data standards. There were approximately 250 comments on this theme. Specific feedback included:

Data Collection

- Adapt data collection systems to county level. A counter perspective noted that local level data analysis can be difficult.
- Obtain line level health data.
- Obtain data on health seeking behaviors.
- Collect and integrate infrastructure data.
- Gather data on conditions beyond COVID-19-related conditions.
- Make data streams flexible enough to allow for one-off types of data collection.
- Understand how data are generated.
- Keep in mind that data are subject to misinterpretation and errors.
- Obtain county level transmission data, vaccination rates, count of new cases, re-infections, admissions and bed use data.
- Create a data repository so that stakeholders can build on what was learned during COVID and continue this work.
- Evaluate what policies need to change to allow for collection of data.
- Ensure there is clear guidance and downward policy support for changes at the state level.

Electronic Medical Records and Healthcare Data

- Implement a system that collects EMR data throughout the country and have a national system in place that interfaces and queries health information exchanges. In the United States there is a very weak connection between the healthcare system, healthcare data and public health data.
- Emphasize healthcare data (especially EMR data) as they are massive valuable assets and often an untapped data source.
- Generate live and aggregated healthcare episode data.
- Create a national patient identifier for patient matching.

Social Determinants of Health Data

- Prioritize and focus on addressing and closing social, racial, economic and structural drivers of inequities. To do this, we need social determinants of health data, for example: race, ethnicity, household crowding, comorbidities, access and socioeconomic status data.
- Assess the conditions that lead to disparate outcomes and the impact of disparate outcomes on health equity.

Geographic Scope

- Keep in mind the importance of geographic scope; this ranges from as local as possible to global data that may offer guidance into new threats.
- Understand how data, including health data, are captured at local levels and what types of analysis/forecasting models should be done at each level; there is utility to forecasting at the zip code level. Get into a virtuous cycle where data modeling benefits state and local individuals. This provides the incentive for state and local individuals to provide the data. This worked well during the Malaria Atlas Project.
- Conduct a landscape analysis at the local/grassroots level to understand gaps, capabilities, goals and needs.
- Educate and collaborate with county governments and other key stakeholders so they are on board and involved in decision-making from the beginning.
- Ensure that data collected from community members gets fed back to the community; otherwise run the risk of resistance in providing data in the future.
- Support local authorities.

Access

- Consider what policies need to be changed to allow access to data needed. There exists a tremendous amount of data, but there are issues accessing it and therefore not using it. A part of access issues is legal barriers including different regulations on standards for data collection among jurisdictions.
- Obtain local data in real time.
- Prioritize complete laboratory and epidemiological data. Often electronic laboratory reporting has incomplete data.
- Focus on better data connectivity.
- Address surveillance data barriers but evaluate guardrails and caveats and how we should be able to access data.

Standards

- Establish a national standard for data reporting including supporting guidelines and resources. Have minimum data sets and definitions. Currently there is no uniform approach to the standardization of data elements and standards vary by jurisdiction.

KEY TAKEAWAYS

A plethora of data exist, but the biggest bottleneck is access to these data. Consider collecting local-level data for guidance into new threats. At the same time ensure that data get fed back to communities, so communities understand the value of providing data. Evaluate what policies need to be changed to allow for increased access and data connectivity. Data must be collected on variables related to social determinants of health to close the gaps in drivers of health inequities. A national standard for data reporting is needed, including minimum data sets and definitions to improve data access and sharing.

REAL-TIME AND GRANULAR DATA

Listening session participants emphasized obtaining granular data, disaggregated data and data at the state and local level that can be collected in real time. There were approximately 70 comments related to this theme. Some participants felt that data become less meaningful when in aggregate and that a flat system where everyone has access to real-time raw data is important. Other participants reflected that public health reporting streams, which are thought of as real time, are outdated. Additional feedback on this topic included:

- Model human behavior and feedback loops and how they generate and interact with infectious disease dynamics and adoption of practices/restrictions.
- Collect data on human behavior during COVID-19.
- Utilize metagenomic and genomic data to put together clusters of how individual transmission events have been related to each other.
- Implement homomorphic aggregation technology.
- Obtain mobility travel data at the local level (currently only available at the national level).
- Prioritize low latency and reliable data reporting. Real-time data have had long lag times from collection to actual reporting, so what is real-time may be outdated. There is currently hesitation with reporting preliminary data that leads to data latency and revisions. It may be easier to report data without revisions, but we can use data set with revisions more effectively. The traditional reporting process causes a loss of granularity.
- Prioritize timeliness over cleanliness and accuracy; corrections can be applied retrospectively.
- Provide metadata about timing and nature of updates.
- Build infrastructure and modeling system such that it considers early noise and bias.

KEY TAKEAWAYS

Obtain disaggregated and state- and local-level data. Include human behavior and feedback loops into modeling for infectious disease dynamics. Metagenomic and genomic data can be used to create clusters to better understand the relationship of individual transmissions. Be willing to report preliminary data without revisions to reduce latency and use data with revisions more effectively. Build modeling systems to include early noise and bias.

COVID-19 AND OTHER DISEASES

There were approximately 70 comments related to COVID-19 and more broadly other infectious diseases, current and future novel diseases. The context of the summary below relates to the scope the Center should consider as high priority. Note that better understanding of transmissibility of infectious diseases was highlighted across these comments.

COVID-19

- Analyze periods of surges.
- Examine COVID-related health conditions and long-term impacts of COVID-19.
- Understand what policies have been effective in slowing down transmission, minimizing deaths and hospitalizations, and improving equity around COVID-19.
- Consider how current COVID data impacts future forecasts on COVID.
- Study the relationship between vaccination and natural immunity against COVID and transmissibility.

Other diseases

- Work with academia and the private sector to understand vector-borne diseases and other zoonotic diseases.
- Focus on noninfectious diseases and conditions as well as infectious diseases.
- Investigate epidemiological-related questions such as transmissibility, disparities in severity and progression of illness, to enhance understanding of how infectious diseases move through communities.
- Maintain advances in pooled testing, which allows for more people to be tested quickly using fewer testing resources.
- Analyze the impact from noncompliant individuals (e.g., mask wearing).
- Evaluate transmission changes and areas of high transmission.

KEY TAKEAWAYS

Analyze COVID-19-related policies that have been effective in minimizing transmissibility, severe illness, death and inequities, and apply these policies to other infectious and noninfectious diseases. Consider how to address challenges of noncompliant individuals. Emphasize cross-sector collaboration to understand vector-borne and other zoonotic diseases. Maintain pooled testing capacity.

SURVEILLANCE

There were approximately 20 comments on the topic of surveillance. Participants noted that it is a bold goal to have a national syndromic surveillance platform and the current state of funding and resources remains challenging to achieve this goal while trying to build up surveillance. People who are conducting surveillance should be more familiar with surveillance techniques and the provision of data needed for forecasting.

KEY TAKEAWAYS

Integrate routine surveillance in the United States and globally, including genetic, serological, active and sentinel surveillance to help with case follow-up. Integrate modeling into existing surveillance programs.

PROPOSED FUNCTION

Connect

PREDICT

- Modeling and Forecasting
- Infrastructure
- Data Collection and Accessibility
- Real-time and Granular Data
- COVID-19 and Other Diseases
- Surveillance

CONNECT

- Partnerships
- Data Sharing

INFORM

- Communications

PARTNERSHIPS

There were approximately 60 comments related to partnerships. Partnership for the purpose of developing a plan for the Center means evaluating and using existing relationships among those working in modeling, policy and public health and learning from each other to inform future work and allow for increased access to data. Recommendations included having a systematic approach to look for peer organizations across government and having structured relationships that focus on addressing questions together. Other comments included:

- Build trust and sustained relationships through ongoing efforts, not solely in a crisis.
- Facilitate a nationwide collaborative with key members; for example, discuss ways health departments, hospital associations and other regional systems share data.
- Leverage partnerships in academia and the private sector, for example, pharmaceuticals, to examine data and diseases and bring faster capabilities.
- Work with local communities to identify critical health priorities.
- Have an advisory committee that represents diverse perspectives; implement one or two collaborations with state and local groups to be used as a model.
- Collaborate on a global scale to share data.
- Encourage stakeholders, including users, to talk to modelers and forecasters to learn from each other's perspectives so that they can inform one another going forward.

KEY TAKEAWAYS

Implement a systematic approach to partnering with peer organizations and having structured relationships. Maintain these relationships outside of "peace" times. Implement a nationwide collaborative and/or advisory committee with key stakeholders.

DATA SHARING

Participants across sessions underscored the importance data interoperability among key stakeholders. There were approximately 30 comments received on this topic. Feedback included:

- Focus on data confidentiality and sharing the data with stakeholders. In the United States there is a very weak connection between the healthcare system, healthcare data and public health data.
- Standardize data use agreements and provide clear guidance to improve efficiency of data sharing.
- Integrate platforms for data curating and sharing.
- Address barriers to data provenance and sharing, such as legal and privacy issues. Regulations are outdated and limit researchers and others who work with data at multiple levels (and hinder the ability to share). This is corroborated in the 2015 Van Panhuis et al. article "A Systematic Review of Barriers to Data Sharing in Public Health": "A centralized mechanism such as a commission or secretariat should monitor, mediate, and facilitate data sharing among various stakeholders to ensure a fair and efficient use of data for the advancement of population health."[4]
- Understand data governance and ownership for better data connectivity.
- Recommend implementing an API that all providers/holders of EMR data will agree to that can be queried. We should put in place a national system that interfaces to the health information exchanges and queries them.
- Encourage more discussion between health departments, hospital associations and other regional systems to share data.
- Employ data governance staff as part of public health departments.

Van Panhuis, et al. discussed the importance of data sharing, but recognized the limitations citing that no systematic framework or global operational guidelines have been created for data sharing in public health.[5]

KEY TAKEAWAYS

Standardize data use agreements, address legal and privacy issues related to data sharing and integrate platforms for ease of data sharing.

4 Van Panhuis, Willem G, et al. "A Systematic Review of Barriers to Data Sharing in Public Health." BMC Public Health, vol. 14, no. 1, 2014, doi:10.1186/1471-2458-14-1144.

5 Ibid.

PROPOSED FUNCTION

Inform

PREDICT

- Modeling and Forecasting
- Infrastructure
- Data Collection and Accessibility
- Real-time and Granular Data
- COVID-19 and Other Diseases
- Surveillance

CONNECT

- Partnerships
- Data Sharing

INFORM

- Communications

COMMUNICATIONS

Listening session participants across all sectors and roles stressed the importance of communication, with approximately 130 comments received related to this theme. Communication in this context is effectively conveying modeling and forecasting results to key stakeholders including decision makers and the public. It includes messaging, visualization, translation, the feedback loop and political considerations.

Participants emphasized that building relationships and maintaining good communication during "peace" times are critical and will make the public savvier consumers. Ongoing communication between decision makers and data modelers could be achieved by having regular meetings where updates and feedback are provided, having dialogue between various levels of public health to ask questions and interpret results, and establishing channels of communication at the outset. It is important to communicate not only the results, but the process of getting to the results including assumptions and uncertainties, as noted in the "Modeling & Forecasting" section. A dedicated communications team working at the interface of modeling, decision-making, and policy-making to ask the important questions and help with interpretation is needed to support this function. To that end, specific feedback related to communications included:

- Translate information from models into something meaningful and relevant to decision makers and communities as processes are applied across public health and data analytics. It is critical to have basic graphics and visuals for the public and translation of analytical results for decision makers.
  - Prioritize effective visualization and communication as they are key components of modeling.
  - Communicate outputs visually and determine what works to communicate complicated data to decision makers.
- Take data and science and translate into action; move from basic to applied research. There is currently a lag in the time it takes to put new science into action; focus on translation includes taking bench research to the public. This should also be part of educational campaigns and translation at the grassroots level.
- Involve community members in translating the data and educating the local public. Hyperlocal relevance and actionability for communities.
- Include community translational scientists and applied researchers, especially when working with modelers on communication and messaging.

- Recommend that CDC creates and hosts a journal (*Journal for Epidemiological Forecasting*, for example) that would build credibility and be useful to multiple stakeholders and not solely the research community (these are different communities).
- Track the audit or provenance of data, ensure it flows through to the end of the consumer.
- Focus on user-friendly tools and communication vehicles.
- Recommend getting faster consensus on key communication points.
- Support ongoing dialogue between people who are doing surveillance and those that are on the front lines.
- Communicate biases in metadata and to modelers. Biases in data should be included in metadata and must be communicated to modelers.

Another premise that emerged from discussions about communication is the politization of data. Participants recommended considering who is asking the question and information regarding their driving interest. Over the course of the pandemic, public health officials have become politically disenfranchised.

KEY TAKEAWAYS

Communications should be a key priority; dialogue among stakeholders and between public health is critical, including communications on processes and any uncertainty. Establish dedicated roles for communicators to interface between data modelers, decision makers and other stakeholders. Effective communication to the public is critical; provide basic graphics and visuals to the public. Involve local community members in the translation of the data for local consumption. Keep in mind the potential politization of data. Be mindful of who is asking the questions and why. Create and host a peer-reviewed journal.

Summary responses to additional questions

1. Listening session participants identified the following new and/or innovate data sources the Center should consider to achieve its goals:
   - Claims and EMR data
   - Google search and mobility data
   - Anonymized Web search queries and location visits by zip code
   - Exploration of Bluetooth data
   - Continuous online surveys (e.g. self-reporting on symptoms, behaviors, beliefs etc.)
   - Pharmacy retail purchases
   - Freely available synthetic data derived from individual level public health data
   - Electronic Test Orders to support active surveillance
   - Diagnostic test results—in home or at urgent-care level
   - CDC COVID-19 County Transmission Level Data
   - Contact network relationships, such as schools and households—ability to track or do syndromic surveillance within schools and households and the connections between them (prior to contact tracing)
   - Exposure notification/proximity data
   - Social services data
   - Mental health data in an ongoing way—measure psychological events; tend to be measured outside of real-time impacts; what are our public health responses to these events
   - High-geographic and temporal resolution real-time data about genomic make-up of new and emerging strains
   - School absenteeism needs improvement
   - Local emergency responder data, ambulance and law enforcement
   - Environmental quality in homes
   - Air/water quality, rain, and precipitation
   - Wastewater

2. Listening session participants identified stakeholders critical in the implementation of solutions in the following categories: pharmacy/lab/health care, academia, science and data experts, public health, policy makers, government, private partners, public and community partners, data suppliers and communications:

- Pharmacy, Lab, Health Care
  - Pharmaceuticals
  - Hospital associations
  - Clinicians
  - Emergency room clinicians
  - Urgent care networks
  - Owners/ custodians of EMRs
  - EMR/EHRs
  - Hospital administrators
  - Healthcare organizations
  - Long-term care agencies, skilled nursing facilities—they are usually left behind, so have them be in conversation up front
  - Diagnostic (labs, diagnostic devices as close to or at the individual level as possible)
  - Clinical Translational Science Alliances
- Academia
  - Academics
  - Colleges of Public Health, Agriculture and Natural Resources, and Family and Consumer Sciences
  - Academic modeling groups and modeling leads at agencies outside of CDC
  - Academic departments responsible for training relevant workforce (Public Health, Comp Sci, Statistics, Biostatistics, Epidemiology etc.)
  - The National Land Grant University Cooperative Extension System
- Science and Data Experts
  - Scientists (modelers, immunologists, virologists, behavioral scientists, epidemiologists)
  - Modelers, regarding methods, assumptions, interpretation
  - Behavioral scientists
  - Data scientists
- Public Health
  - Public health officials
  - STLT Health departments, and those supplying data sets of interest—there often is not a good understanding of the metadata about the data sets themselves
  - Health department staff
  - Health departments to describe data generation processes and limitations
  - Public health services groups within each state (public health laboratory, local health department, state health agency) and groups such as organizations that are responsible for managing a centralized information technology services group
  - Local public health officials with meaningful audience/linkage with governing officials/ politicians in power
  - Major epidemiologic and public health organizations, e.g. APHA, ASTHO, CSTE, ASPPH, AMA, ACLA

- Policy Makers
  - Political leaders
  - Policy makers
  - Politicians to frame key questions
  - Decision makers, such as health officers and mayors
  - Elected officials, including state, local and federal health officials
- Government
  - Other federal government agencies who would be interested in the outputs of the models
  - Federal health authorities
  - CDC, other federal partners
  - Make sure there is coordination across federal entities about what needs to be collected as well as validating the ability to collect those data and determine within what system are those collected in
  - Biodefense space/DOD
  - Funding agencies
- Private Partners
  - Cloud providers (Amazon, Google etc.)
  - Private sector modeling groups and data providers
  - Private companies/employers—have interest in workforce, supply chain etc.
  - Employers—businesses (can implement workforce policy) (important to supply chain)
  - Information technology vendors
  - Big tech and data firms
  - Liaisons—those developing public/private partnerships on both sides
  - Funding agencies
- Public and Community Partners
  - Education organizations—critical for influencing behavior
  - Range of end users at community level—outside PH that rely on timely, accurate data
  - Community groups/activist groups
  - The lay public—on the behavioral side
  - Communities that experience disparately large burdens—and more severe outcomes
  - The public—specifically communities experiencing disparate impact and high burden
  - Involve churches and religious leaders since they provide lots of support; Ministries of Health—without their support nothing can be done
  - Community decision makers must give support and training and help them understand what you are doing
  - Start with customers/consumers
- Data Suppliers
  - Data warehouse companies and other data generators
  - Suppliers of the raw input/data—whoever owns or has access to the data that we need
  - Coroners/morticians. Fatality management—coroners or anyone who handles fatalities
- Communication
  - Communication specialists
  - Media

3. Listening session participants identified specific staff needed in the implementation of solutions:
    - Informaticians at every level
    - IT experts at every level
    - Technologists
    - Machine learning experts
    - Software engineers
    - Data collectors
    - Data scientists
    - Data analysts
    - Statisticians
    - Data use agreement experts
    - Subject matter experts from the areas under consideration, e.g., infection preventionists, microbiologists
    - State and local health departments
    - Policy, legal, and ethics expertise that can advise on policy levels and aspects
    - Software engineers
    - Communications/PIO
    - Public health data subject matter experts

APPENDIX A

List of Participants

DMI LISTENING SESSION 1
THURSDAY, JULY 22, 2021

Hosts and Observers

Judy Monroe, *CDC Foundation*
Daniel Jernigan, *Centers for Disease Control and Prevention*
Brandon Talley, *CDC Foundation*
Corinne Graffunder, *CDC Foundation*
Sherri Berger, *Centers for Disease Control and Prevention*
Alison Kelly, *Centers for Disease Control and Prevention*
Marc Lipsitch, *Centers for Disease Control and Prevention*
Caitlin Rivers, *Johns Hopkins University*
George Dylan, *Ginkgo*

Facilitators and Notetakers

Neha Agrawal, *Illinois Public Health Institute*
Laurie Call, *Illinois Public Health Institute*
Sydney Edmond, *Illinois Public Health Institute*
Sarah Gillen, *Public Health Consultant*
Samantha Lasky, *Illinois Public Health Institute*
Susan Martinez, *Illinois Public Health Institute*
Melissa Moorehead, *Public Health Consultant*
Lucy Peterson, *Illinois Public Health Institute*

Meeting Participants

Allisen Penn	University of Georgia Extension	Academia	Practice
Andy Weber	Council on Strategic Risks	NGO	Policy
Chandra Ford	University of California, Los Angeles	Academia	Research
Don Burke	University of Pittsburgh	Public Health	Research
Evgeniy Gabrilovich	Google Health	Private Sector	Research
Janet Hamilton	Council of State and Territorial Epidemiologists	Public Health	Epidemiology/Surveillance
John Zicker	Kinsa	Public Health	Forecasting
Justin Greenberg	Kinsa	Public Health	Forecasting
Karen Smith	Google Health	Private Sector	Research
Michelle Meigs	Association of Public Health Laboratories	Public Health	Epidemiology/Surveillance
Nick Reich	University of Massachusetts	Academia	Forecasting
Ryan Tibshirani	Carnegie Mellon University	Academia	Forecasting
Sargis Pogosjans	Public Health—Seattle and King County	Public Health	Epidemiology/Surveillance
Scott Becker	Association of Public Health Laboratories	Public Health	Epidemiology/Surveillance
Ather Javaid	AdventHealth	Health Care	Epidemiology/Surveillance

DMI LISTENING SESSION 2
WEDNESDAY, JULY 28, 2021

Hosts and Observers

Judy Monroe, *CDC Foundation*

Daniel Jernigan, *Centers for Disease Control and Prevention*

Alison Kelly, *Centers for Disease Control and Prevention*

Brandon Talley, *CDC Foundation*

Corinne Graffunder, *CDC Foundation*

Facilitators and Notetakers

Neha Agrawal, *Illinois Public Health Institute*

Elissa Bassler, *Illinois Public Health Institute*

Laurie Call, *Illinois Public Health Institute*

Sydney Edmond, *Illinois Public Health Institute*

Sarah Gillen, *Public Health Consultant*

Samantha Lasky, *Illinois Public Health Institute*

Susan Martinez, *Illinois Public Health Institute*

Melissa Moorehead, *Public Health Consultant*

Lucy Peterson, *Illinois Public Health Institute*

Meeting Participants

Name	Organization	Sector	Role
Amanda Beaudoin	Minnesota Department of Health	Public Health	Epidemiology/Surveillance
Annie Fine	Council of State and Territorial Epidemiologists	Public Health	Policy
Bob Davis	Indiana Department of Health	Public Health	Healthcare IT
Bryan Grenfell	Princeton University	Academia	Research
Cecile Viboud	National Health Institute (NIH)	Other	Forecasting
Charlotte Jessica Metcalf	Princeton University	Academia	Research
Jeffrey Shaman	Columbia University	Academia	Research
Justin Lessler	University of North Carolina Chapel Hill	Academia	Research
Kim Vu	New York City (NYC) Department of Health and Mental Hygiene	Public Health	Other
Ronald Ginder	Indiana Department of Health	Public Health	Healthcare IT
Roni Rosenfeld	Carnegie Mellon University	Academia	Forecasting
Ruth Lynfield	Minnesota Department of Health	Public Health	Epidemiology/Surveillance
Salima Kasymova	Howard University	Academia	Research
Sharon Greene	NYC Department of Health and Mental Hygiene	Public Health	Epidemiology/Surveillance
Steven Riley	Imperial College	Academia	Research
Tara Rhone	AdventHealth	Health Care	Epidemiology/Surveillance
Virginia “Ginny” Pitzer	Yale University	Academia	Research

APPENDIX B

Methodology for Listening Sessions and Qualitative/Thematic Analysis

The two video-conference-based listening sessions were three hours each and held on July 22 and 28, 2021. Attendance across both listening sessions was 41 participants. The breakdown among participants providing input, observers, and staff support was as follows:

July 22

Total Attendance	**21**
Participant Attendance	16
Observer Attendance*	5
Facilitators and Notetakers	8

July 28

Total Attendance	**20**
Participant Attendance	17
Observer Attendance*	3
Facilitators and Notetakers	8

**Note that not all observers attended the breakout groups.*

Participants contributing to the listening sessions spanned a variety of sectors and roles, with an emphasis on the academia and public health sectors, and epidemiology/surveillance and research roles, as shown in the table below:

July 22

Primary Sector	
Academia	4
NGO	1
Health Care	1
Public Health	5
Private Sector	1
Other	0
Primary Role	
Epi/Surveillance	4
Forecasting	3
Practice	1
Research	3
Healthcare IT	0
Policy	0

July 28

Primary Sector	
Academia	8
NGO	0
Health Care	1
Public Health	7
Private Sector	0
Other	1
Primary Role	
Epi/Surveillance	4
Forecasting	2
Practice	0
Research	7
Healthcare IT	2
Policy	1

The high-level agenda for both listening sessions included:

- Welcome and Introductions (Laurie Call, IPHI)
- Level Setting: Background and Context (Dr. Judy Monroe, CDC Foundation and Dr. Daniel Jernigan, CDC)
- Public Health Data and Research Goals and Priorities for the Next 5 Years (Large group session)
- Methodological and Research Needs (Breakout groups)
- Priorities, Strategies, and Solutions to Support Utilization and Implementation (Breakout groups)
- Closing (Laurie Call, IPHI)

For each session, feedback was first collected in a large group discussion. Participants were then divided into breakout groups with a trained facilitator and notetaker. Each breakout group was assigned specific questions to respond to via group discussion and an interactive online whiteboard tool. The facilitators solicited participant responses to the questions while notetakers took notes in real time on the screen. Finally, the participants were invited to contribute post-meeting feedback through comments and additions to the session notes, though there were few contributions to this component. After Listening Session 1, facilitators and notetakers debriefed and made slight modifications to the questions to improve participant understanding and engagement. Explanations about how the questions changed, if at all, for Listening Session 2 are written as subtext within Appendix C, which contains the detailed questions and notes captured during both listening sessions. During the qualitative analysis, participant responses were counted as a single comment though the comment often had multiple components and/or addressed multiple themes.

An IPHI staff member completed the qualitative analysis. The raw notes were reviewed to identify themes and comments were organized within the themes. Staff identified recommendations based on concurring and consensus comments. Interrater reliability did not apply, rather data analysis included a count of the number of times a similar comment or recommendation was made. An approximate number of counts for each theme is listed in this report.

APPENDIX C

Listening Session Technical Notes

DMI LISTENING SESSION 1
THURSDAY, JULY 22, 2021

MAIN SESSION

1. What are essential or priority questions for analysis in the shorter term (1–3 years)?
    - Priorities are funding dependent; public health teams are trying to work with academic partners because data resources aren't necessarily available at the public health agency, especially the county level
        - Difficult to analyze data at county level
        - Focusing on climate-based events and forecasting (e.g. heatwaves, forest fires, etc.)—which communities will be affected
        - Working to identify resources and analyze data in the future to identify which parts of the county are most susceptible
        - Leveraging partnerships in academia or the private sector to examine vector-borne infectious diseases and other emerging diseases that are animal centric or zoonotic in nature
        - Building up respiratory surveillance, funding remains challenging especially at a local level where research projects require resources beyond local capacity
    - Probing question: If money and resources are available, what are the ideal priorities?
        - Challenge working with communities is identifying critical health priorities county-by-county
            - One participant and their team are working with community health coalitions to map priorities—connecting with departments of health and other researchers
            - Develop methods to map priorities to future initiatives
            - Focus on working together collaboratively instead of competitively, the need to be part of a larger data center
        - Forecasts come through and are evaluated, but often they aren't as robust as desired
        - Key challenge: matching questions that have direct public health relevance and models that can be used to drive public health decisions
            - Which models can be used to make useful predictions and have accuracy?
        - Finding overlap of what public health practitioners need and what models can tell us requires careful consideration and examination of what data we have and what models can do
    - One participant and their team created four major data sets of different kinds with international data (e.g. examined visits to different community centers such as hospitals, stores, etc.); data was then used for forecasting and analysis of observance of social distancing
        - Examined surge volumes and different COVID conditions

- Would be helpful to adapt these data collection systems to county level in addition to national level
- Make data collection move beyond just COVID-related conditions
- This participant and their team have new focus on vaccine equity—quantified what questions users have about vaccinations; made data available to public health professionals to improve educational and outreach programming
- Recommend working with external stakeholders and collect what data they could find useful in their respective fields

- Recent data sets and dashboards from Google:
  - COVID-19 Community Mobility Reports: https://www.google.com/covid19/mobility/
  - COVID-19 Symptoms Search Trends: https://pair-code.github.io/covid19_symptom_dataset/?country=US
  - COVID-19 Vaccination Access Dataset: http://goo.gle/covid19vaccinationaccessdataset
  - COVID-19 Vaccine Search Insights: https://google-research.github.io/vaccination-search-insights/
- Building off the comment about forecasts—the way we forecast the problem must be directly related to the question we are trying to answer; translations must also be made as we apply this process across public health and data analytics; needs to be a continuous process
  - Challenge of pandemic forecasting: forecasting isn't always in line with what is needed by sectors in question
  - Process must be continuous, and work must be completed on a spectrum with consistent communication
  - Important consideration: whether or not the data we are using is actually measuring what we are hoping for—are the mathematical models being used actually appropriate for the forecasting we are doing?
  - Data is subject to errors and misinterpretation—needs to be considered in ongoing discussion with people who work on data collection and forecasting
- Another area to think about is on the "analysis" side rather than forecasting...we have a real need to understand what public health policies have actually worked to slow transmission/decrease deaths and hospitalizations/increase equity in COVID...forecasting is important, but it is even more useful when we know what to do with the results...these findings for COVID will be critical for other public health events. This kind of analysis is complex and multi-sectoral
- Support for comment about some targets of forecast being unclear and maybe not that informative. One way to structure the discussion could be for decision makers to define some paradigmatic decisions and as much as possible what they are trying to optimize, then back and forth with the forecasters to elicit understanding of the goal and the target. This is not a one-and-done process.

  a. Are they different over a longer time horizon (3+ years)?
  - An ongoing consideration in both short- and long-term is the feedback loop that happens; effort to relate end intervention and action with the forecast will take several cycles to perfect
    - This is a key challenge that relates to why forecasting can be disappointing
  - Challenge with forecasting is having good ongoing inputs about the people—this is often considered in epidemiology
    - Mobility pieces are key; there hasn't been granular details presented about where people are moving AND how they move
    - Consider not just movement in the general population but also mobility and utilization pieces in health care as well (e.g. movement between acute to long-term-care facilities, hospitals, etc.)

        - Granular information can also be used when there is a specific disease or condition of concern
            - For different diseases and conditions, there are different things that are important
        - We miss the gap on granular understanding of how people are moving in the population and responding as these challenges and diseases arise
    - High-level questions of organization and structure—modeling has many components (data identification, analytics, computational modeling, economic interpretation, feedback, etc.)—points to a bigger question about how we conceptualize the process
        - At other organizations, this is left to the ecosystem to evolve itself—people are expected to come forward and present their ideas
        - There are only so many niches in the infectious disease and biological modeling ecosystem
        - So, does the Center conceptualize itself as having these components? Will we cultivate groups by disease or components (social or genetics)? Will we organize ourselves around these components?
        - Some level of structure would be needed
        - Challenge: modeling infectious diseases is easy, whereas modeling humans/human activity is more challenging—what are the boundaries that the modeling center will do?
            - If we limit ourselves to infectious diseases, we might not get very far
            - Recommend taking on the harder task of analyzing human activity and feedback loops and how they generate and interact with infectious disease dynamics

2. What would be bold, meaningful, and feasible to see accomplished in the next 5 years in the following categories?
   Research (comments pulled from question 1):
    - Focusing on climate-based events and forecasting (e.g. heatwaves, forest fires, etc.)—which communities will be affected
    - Working to identify resources and analyze data in the future to identify which parts of the county are most susceptible
    - Working to also leverage partnerships in academia or the private sector to examine vector-borne infectious diseases and other emerging diseases that are animal centric or zoonotic in nature
    - Building up respiratory surveillance, funding aspect remains challenging, especially at a local level where research projects require resources beyond local capacity
    - Working with community health coalitions to map priorities—connecting with departments of health and other researchers
    - Using these priorities to map out future initiatives and for predicting is challenging
    - Matching questions that have direct public health relevance and what models can be used to drive public health decisions
        - Which of these models can be used to make useful predictions and have accuracy?
    - Finding overlap of what public health practitioners need and what models can tell us requires careful consideration and examination of what data we have and what models can do
    - Recent data sets and dashboards from Google (these can be adapted for stakeholders and partners):
        - COVID-19 Community Mobility Reports: https://www.google.com/covid19/mobility/
        - COVID-19 Symptoms Search Trends: https://pair-code.github.io/covid19_symptom_dataset/?country=US
        - COVID-19 Vaccination Access Dataset: http://goo.gle/covid19vaccinationaccessdataset
        - COVID-19 Vaccine Search Insights: https://google-research.github.io/vaccination-search-insights/

- If we limit ourselves to infectious diseases, we might not get very far
- Recommend taking on the harder task of analyzing human activity and feedback loops and how they generate and interact with infectious disease dynamics

Data Gaps:

- In addition to the variety of data, geographic scope matters—ranges from as local as possible and insight into global data that may offer guidance into the new threats—applies to all data we are trying to obtain
- Metagenomic data
- Granular data collected at the state level (data given to the CDC is usually less granular)—synthetic data sets generated on state data should be made freely available; avoid tracking individuals and respecting privacy; work around data use agreements
- Would others agree on the value of a large table that by county/age/possibly race/ethnicity includes the numbers of persons with various comorbidities and their overlaps (combos of 2, 3, more)? This was a big challenge in vaccine planning for both 2009 H1N1 and COVID-19
- Race/ethnicity and other intersectional data is missing nationally
- Urgent need for us to bring together better public health data that includes more information about the populations, including vulnerabilities, social services, etc.; put this together with healthcare data that can be linked—will lead to effective analysis and forecasting
- Obtaining granular data is important; we should also gather data on the structural drivers of health inequities and health outcomes—collect information on multiple indicators to support work on SDOH

Implementation Factors (i.e. policy aspects, system, structural, workforce, etc.):

- National syndromic surveillance platform as a bold goal
- Legal reporting requirements for social determinants data—when there isn't a legal requirement, that causes data gaps; during COVID, there have been challenges and concerns around providing data for certain populations that's not available; labs aren't set up to collect this information (infrastructure is absent because there is no requirement to report or funding); long-term goal, need national or higher level requirements on reporting (varies state-by-state); levels of detail do fall off the more the data is aggregated
- The public health system needs more and more robust informatics workforce.
- Support for data coming from health care—claims and EMR data can be extremely valuable and timely sources of information and linked and combined with other important features. Not only serving as early indicators of activity, but alternative measures of activity and alternative measures of ground truth entirely
- Biggest implementation factor in academia is that incentives of doing this kind of work (building real-time models meant to be operationally tracking real-time infectious disease) doesn't exist; if we are relying on this work to be done in academia, challenges will arise in that academics will work toward publishing; the type of work that's most effective isn't always work that should be turned into paper—often getting into weeds and details of problem doesn't fall into traditional researching and modeling (doesn't mean this work shouldn't be done—just needs to be considered when discussing incentives)
- The same can come from private sector—novel data sources can increase timeliness of forecasting and reporting (e.g. Kinsa is communicating with underserved communities and sentinel groups very early in the illness cycle)
- Policies around data provenance and usage (especially in the context of big mixed data sets from various sources/data lakes)
- At a very basic level how is health data captured (rural, migrant, and underserved populations)

- I think public health departments need to build up workforce—for many jurisdictions it falls under the general "epi" groups who are not always able or trained to do this work. Dedicated funding for modelers would be best
- A critical aspect of developing policy based on models is how community partners can be involved in educating the general public as part of a continuous loop
- Supporting training programs that merge public health data challenges with advanced statistical modeling and analytical techniques is critical to help the field grow

3. How should infectious and noninfectious disease modeling research and application be prioritized and enhanced? Are there lessons from one that can be applied to the other?
    - We need to do this work beyond infectious disease (facilitator noted as earlier priority)
    - Must acknowledge that if we are working toward long-standing social inequities, that there needs to be specific funding and focus on conditions, diseases (infectious or not), environmental factors, etc., and prioritize cross-sectional areas of research that won't further divide or silo the social determinants we've currently identified
        - Much of current research identifies the conditions that are exacerbated by the factors and communities that are most at risk (e.g. climate change and infectious disease—for example, we've seen the effects of flooding on fecal oral diseases that will particularly effect folks who are experiencing homelessness)
        - Prioritization and focus needs to go toward closing social, racial, and economic inequities that are part of public health's aim to address racism as a crisis
    - There is danger if the Center focuses too narrowly on infectious disease—will possibly preclude group from completing its mission
        - Having systematic effort to look for peer organizations across the government and having structured relationships that focus on how we address this question together is a good place to start
    - There are many complex contagions and there's a developing field of rumors, beliefs, polarization, etc., that is impacting modeling—there's much that can be done in this space
        - Data has become politicized
    - Between both infectious and noninfectious diseases, supply chain issues and where critical constraints on the supply chain exist are challenging and often misunderstood by public health
        - In the pandemic, there was overwhelming of the infrastructure (this is also true of influenza outbreaks)
        - There is lack of understanding of these issues, and we need to consider which things we are looking at and what the key issues are
        - Medical supply chain and how things move—consideration is critical
        - How people get their information is a huge challenge that impacts behavior change and progress of diseases and conditions—not yet well understood in public health

4. What new or innovative data sources should be considered?
    - Wastewater
    - Claims and EMR data
    - Pharmacy retail purchases
    - Continuous online surveys (e.g. self-reporting on symptoms, behaviors, beliefs, etc.)
    - Social services data
    - Google search and mobility data
    - Electronic Test Orders to support active surveillance
    - Freely available synthetic data derived from individual level public health data
    - Diagnostic test results—in home or at urgent care level

- One participant and team specialize in producing novel signals by aggregating (in a privacy-preserving way) the data that users generate while interacting with different Google products and services, such as anonymized Web search queries or location visits. They make their insights and data available at zip code/county level. They hope their data can serve as a free and near-real-time proxy for conducting surveys or collecting data from healthcare systems, both of which can be expensive and time consuming.
- CDC COVID-19 County Transmission Level Data
- Mental health data in an ongoing way—measure psychological events; tend to be measured outside of real-time impacts; what are our public health responses to these events
  - +1 from another participant
- Contact network relationships such as schools and households—ability to track or do syndromic surveillance within schools and households and the connections between them (prior to contact tracing)
- Air/water quality, rain, and precipitation, etc.
- I think the Bluetooth data should be explored
- High-geographic and temporal resolution real-time data about genomic make-up of new and emerging strains (maybe this is what was meant by metagenomic data?)
- School absenteeism—needs improvement
- Local emergency responder data, ambulance and law enforcement
- Exposure notification/proximity data
- Environmental quality in homes

5. What is most important in the identification and prioritization of research and infrastructure development?
   - We need to understand what types of analysis/forecasting should be done at what level e.g. local, state feds, government vs academia vs private sector
   - Most important is who is asking the question—are they interested in cases, dollars, or votes?
   - Identifying targets that have both public health utility and are tractable from a modeling perspective.
   - For infrastructure development—focusing on getting data from health care to public health in a timely manner is important…
   - …And having the right people on the public health side to do something with that data
   - Reliable, low latency, local public health data reporting infrastructure. Standards for data reporting
   - Ensuring that a diversity of modeling approaches are represented in the pool of models solicited/curated/considered
   - Community needs assessments
   - Promote convergence of clinical and public health data
   - Evaluation of the current gaps within local/state/federal departments
   - Another participant is doing great work on their hub. This commentor would love to capture lessons learned and sharing of that from forecasting teams
   - Funding for sample collection and analysis
   - Identifying what policy needs to be changed to allow collection of the data we need
   - Ongoing dialogue between stakeholders, data processors/curators, and modelers

BREAKOUT SESSION 1

6. What are the data that would be most useful for improving modeling and forecasting?
    - Incorporate SSVI and CCVI into modeling in some form
    - County Level Transmission Data, vaccination rates, count of new cases, re-infections, admissions, bed use
    - Disaggregated data (line/individual level data)
    - How to answer when background isn't in modeling, but want to be sure answers in breakout group reflect constituents (informatician/public health). What do we mean by larger sources of data?
        - e.g. case reporting from clinical care? Lab reporting?
        - How to make sure we are not double counting?
        - Challenge: obtaining data sets from the same person would need a data lake, technology where you can marry different data sets. How do we ensure data is representative, would it be helpful to disaggregate?
    - Some data is inaccessible, whether due to inaccessibility, data governance concerns, siloes, etc. Causes concerns with how to aggregate, understand data, does it mesh with the purpose/use for the data? Useful to understand data governance and ownership, better data connectivity.
    - What influence does policy have? We have a lot of data but can't use due to legal hoops.
    - Clarifying question: what are we modeling and forecasting? Are we meant to consider any of priority topics?
    - There is a large dichotomy that affects answer: modeling and forecasting in pandemic versus seasonal epidemic
    - Framing the conversation—Clarifying comments from observer: between CARES and APA, there is over $1 billion available for DMI. Talking about modeling and forecasting—how to spend this money to ensure that the data that are available are best suited to assist with decision-making. Negotiating data agreements could be one, and another is what are big investments we can make to revolutionize epidemic transmission and viral transmission modeling and predictions?
    - In terms of where to put money, build system that links laboratory data and EMR data and various public health systems that is rapidly accessible—consider challenge of COVID.
        - Data streams must be flexible enough to allow for one-off types of data collection, for example, vaccination sites and testing sites
        - Must be able to gather data and link that data
    - We are almost always in a situation of modeling a convolution of things—there isn't enough resolution of the data to make something predictable
        - As an example, if we are modeling COVID cases, it would depend on if you are modeling for prisons or nursing homes. It's hard to project count forward in time unless we know what it is composed of. Convolution of different processes happens.
        - Second point is do more with latency and revisions; hesitation with reporting preliminary data. That leads to latency. That should be a modeling question, not on the data release folks. The degree of noise, etc., should be part of the modeling team.
            - Summary: greater resolution, more metadata, more fields in the data, and releasing data sooner
        - Public health reporting streams that we think are real time are actually outdated
        - Example of COVID case data: when COVID cases are reported to CDC, what is being published on county and state websites and what is being made available to the public are COVID cases that are aggregated by report date

        - These cases aren't being aggregated by test date, but rather report date
        - The test date is the more relevant piece of information—it's likely easier to report data without revisions but, if we are willing to produce a data set with revisions, we can use that data more effectively

7. What challenges have you experienced accessing USG and other data for your current research?
    - Standardization of data elements varies depending on jurisdictions since they have their own systems in place
    - Data becomes less meaningful when trying to aggregate and access/compare data to understand data across jurisdictions. Don't have a lot more information than what is available on the national level. Jurisdictions have different regulations on standards for data collection.
    - Do the questions we seek to answer include: what do we have now, and what are we unable to model due to these obstacles?
    - Part of challenge is that data isn't standardized, even remotely
        - There is no one single entity that is able to move quick enough and standardize their data set
        - What is our minimum data set? What are our definitions for each of the components?
        - Needs to be automated—standard frameworks can be determined ahead of time and then adapted to current needs
        - This is a holdup with electronic case reporting (eCR) - determining case definitions can take up to a year. So many other data sets need to be understood across the country.
    - Barriers to state and local jurisdictions to case reporting forms put out by CDC—what are they? (probing question from observer):
        - Data that's coming in is from electronic lab reporting—you're lucky if you know the county of the person. ELR doesn't have data you often need.
            - You only get the name, test date, and lab result. Humans have to manually go back to search for missing information. This kind of stuff needs to happen electronically.
        - Epidemiologists, et al., then must call the lab and ask about the data—if we can't get this data reliably, it isn't sufficient—needs to be accessible electronically and not just via fax or telephone
        - To get the data for good modeling, we must build system and infrastructure (policy and technology)—this doesn't exist yet
        - You can do this eventually, but can you do this in time to make a difference?
        - To get the good data to build modeling we need to get infrastructure in place, including policy and technical infrastructure. Does not exist. But it also needs to be timely and to make a difference.

8. How should modeling be used to inform the design and interpretation of epidemiologic studies and surveillance systems?
    - How does data/observations from COVID impact future observations? What surveillance systems are in place to capture a set of/volume of data? Seattle had a surveillance system in place pre-COVID but, when the pandemic hit, the volume of data broke the system. It's useful to show potential volume of diseases and what kind of diseases
        - Helpful to have a team of statisticians, epidemiologists, etc., to observe
        - Driving funding and persuasion
    - Surveillance data can help with case follow-up, and there are other technologies that can help support that observation to understand what needs to be captured, should it be captured, and what comes in from the source data?
        - Capturing "right" data (place/where it lives) for national databases

    - Potential long-term impacts of COVID, impacts of surveillance (e.g. reporting positives, negatives, etc., to understand landscape)
- Question from observer: One goal of forecasting is to forecast accurately. It can also be a tool to change interventions, prevention strategies, and policies, which can change a forecast's trajectory (if done well). Do we need to be thinking about how to incorporate these goals into modeling systems?
    - Many events can happen between the starting point and the end of the forecasting, including assumptions around interventions and events that can influence modeling. Better coordination with folks with modeling skills, with policy skills, and public health knowledge/care would be needed.
    - Also, there's a need for public health to invest in modeling, private partnerships. Incentives to develop modeling as a core competency within public health departments, as well as placing it as a priority
    - Another participant agrees 100 percent. It would also help to articulate what would be useful for public health departments to people who do not work with public health daily, including data on non-/infectious diseases (e.g. newborn screenings—95 percent tested in a public health lab, but no real models to see what that means for impact on programs)
- It would be helpful for people who work on modeling to be part of discussion of standardization. Decisions that may be innocuous to you are important to modelers.
    - An obsession with forecasting exists in the modeling community and on the consumption side, which overshadows production of real-time estimates of current disease states (i.e. "nowcasting") [agreement voiced from another participant]
    - This is a precursor to forecasting
    - Providing a more accurate picture of situational awareness with nowcasting is more important than forecasting
- Data we get is the tip of the iceberg—it is representative of people who are sick enough to see the doctor (surveillance data is less common). Typically, we only see a select group of patients and don't see majority of people who are affected.
    - The longer the lag between the clinical event and the outcome that leads people to seek care, the harder it is to get appropriate data from all patients.
    - Nowcasting would've been more helpful than forecasting
    - We need more training on modeling, looking at models, and describing what they mean and don't mean
        - Public health professionals, politicians, etc., need to be taught what models mean
    - Nowcasting needs to be better understood and have more prominence
- Part of complication with forecasting has to do with whether or not the forecast is factual or counterfactual. For example, what are the assumptions, are people still wearing masks, will there be more school closures, etc.
    - Choose to model one version of the future predicated on assumptions or you could not to—these are two distinct problems
    - There are very different interpretations of predictions based on whether they are factual or counterfactual. A forecast that does not make these assumptions is different.
    - Problem when we discuss 40 to 50 models being produced is that some of these models should be interpreted as scenario projections—we need to be clear about assumptions and what we are looking at
        - Clouds predictions of forecasts
    - Nowcasting doesn't make statements about the future, rather we are asking what is happening right now (much cleaner and approachable problem)

9. What structure(s) are necessary to support forecasting and modeling to ensure value is added to the existing research agenda and community?
    - Is structure defined anywhere, or is it broad-based, technical, collaborative structures?
    - A national standard is useful to have. Differences/disparities between well-resourced jurisdictions versus less funded/resourced would exist with or without a global pandemic. Seattle is fortunate to have global institutes on modeling, which was helpful for understanding/having attention and funding. Other communities did not have that—would provide value to work. Defining value would also be served by having a national standard.
    - National standard would lead to providing resources, guidelines, etc. (baseline framework). Is the point of a Center to provide a template, point of reference, basic need resources; or would a different approach be relying on local knowledge and partnerships?
    - The con of not having a framework would make collection, modeling messier
    - Would there be anything useful in understanding/having a landscape analysis of public health departments to understand gaps, capabilities, and needs? What are localities actually focusing on? What was capable/what were jurisdictions funded to do (e.g. with CARES Act $$)?
        - Meeting people where they are regarding technical capabilities/expertise—how can tech help support jurisdictions with tech gaps (e.g. capturing data, reporting data, creating feedback loops)?
    - Considering incentives, there seems to be a need to build infrastructure to create an incentive system to get people to participate in real-time efforts beyond what's academically publishable
        - CDC could create/host a journal (Journal of Epi Forecasting, for example) that incentivizes work that would be useful to stakeholders and not necessarily the research community. Those are not the same communities (a second person liked idea of journal).
        - Allow people to create products out of their work that doesn't require them to juggle logistical aspects as in academia
    - Have standardized and ongoing way for researchers and practitioners to discuss these topics
        - Have researcher present what they are interested in looking at and ask what would be useful to public health practitioners—need to create synergy
        - There isn't much understanding on the public health side of what research looks like, especially data analytics and modeling
        - These decisions are important and allow public health professionals and researchers to collectively decide what to do
        - Also important to have health care understand how the data are used for public health purposes
        - Ask the question of researchers: "Do you think [x] thing we want to do is useful to public health?"
    - To date, infectious disease modeling community has been run out of academia but it doesn't have to remain this way
        - Opportunity to realign incentives
    - We do epidemiology all the time, why do we not have same kind of ongoing work in analytics that require new tools
        - This isn't different from epidemiology, it's new tools and new science that can go in a variety of directions
        - Why are public health professionals leaving this to researchers to do this for us?
    - Another participant agrees but feels this cannot happen overnight
        - Huge amount of duplication in forecasting and modeling. Weather forecasting: there are people who work on this academically and then there are those who do it outside of academia, and there are those who communicate about weather to the public—these groups are each unique and different.

    - There is lack of centralization of tools needed to create models and forecasts
    - We often try and solve versions of the same problem starting from scratch—there's so much duplication. CDC is writing code, but all separately.
    - Code sharing and standardization are imperative—the ways forecasts are produced is unproductive and duplicative
  - Centralization/hub is something this center could do—don't suppress innovation but also don't let everyone consistently reinvent the wheel
    - Demonstrates how unsystematic our public health system is
    - There's much to be said about autonomy, but there's also much to be said about coordination
    - Challenge with putting this work at CDC is that this work has the chance of being politicized or suppressed
  - Not all models have performed well historically, so a government organization like CDC is hesitant to show preference to some parties over others—we will need to be mindful of neutrality in this space when selecting models

10. What are key interfaces between modeling/forecasting and genomic surveillance of pathogens? (SKIP if time is short)

    THESE QUESTIONS WERE SKIPPED IN BOTH BREAKOUT GROUPS (BOG) 1 AND 2

11. What are specific recommendations for the integration of verification and validation processes to ensure quality results? (SKIP if time is short)

    THESE QUESTIONS WERE SKIPPED IN BOTH BOG 1 AND BOG 2

12. What structure or design would best achieve the desired scope including the coordination and collaboration with the STLT, academic, health care, private industry, and interagency organizations?
    - There are two components: 1) functional nature 2) governance; not sure if the structure needs to be the same for both. For the functional side, there's a natural flow of processes of data collection, data curation, data analytics, mathematical modeling, agent simulations, economic analyses, visualizations, and communications—these are all components of an effective modeling for decision support. I wouldn't try to superimpose anything about the players onto that functional organization.
      - Then you need some sort of separate governance with participation from those different sectors to make sure that feedback to CDC is there
    - We no longer need to "dumb down" data at every level as it goes up. A flat system where everyone has access to real-time raw data because we now have tools to deal with data in those volumes. The traditional way in which each level up to the federal government decides what they will report causes you to lose a lot of granularity and texture.
    - We need to get into a virtuous cycle where whatever is being provided from a modeling perspective needs to be of benefit to state and local individuals (couldn't be derived on their own). That will provide the incentive for them to provide that data. It will be a lot of work because, regardless of how good we automate the system, there will be a human in the loop that will have to move things along. If we can get into a virtuous cycle where they want to provide that data in some capacity because they get a benefit out of it, then that's what we should do. There are examples where this has worked well:
      - The Malaria Atlas Project—developed maps of malaria incidents in Africa and other parts of the world. Local Ministries of Health could use that to direct malaria efforts. They couldn't generate those results on their own, but once they saw the benefit, they became part of a positive feedback loop. So we should find opportunities where that can be generated to decrease tension between local, state, and federal levels.

    - There must be a benefit, or the efforts won't be sustainable
- +1 on the comment about making sure information gets fed back. From a community basis, they collect a lot of data that gets fed back into a lot of databases; it doesn't always get fed back. Then they run into resistance in communities to continue to give data.
- Maybe we should reverse the questions in terms of what are the key questions that need to be addressed for the public health response at the locality level and then subsequently what are the modeling results that could support that effort (and what are the data that could support that modeling) is another way to support this
    - This will highlight what the local public health officer really needs to do to drive that response. This is the customer that the Center should focus on (mayors, governors, and public health officers at local level).
- There is another feedback loop that is related to that: the value of information along the pathway; that if you're modeling and you have uncertainty in your outcome, then one of the ways to minimize the uncertainty is to have additional data that is better, different, or like data. That in turn can inform what data needs to be collected and how much it will cost. You can use that to make decisions about research agenda.
    - If I collect this additional data, will it give me value in my subsequent decisions and narrow my confidence bounds about cost?
    - This is another type of system feedback
- The federal government should pay for data. Whether that's paying for samples or paying a company to take those samples and create data streams, that's what's lacking and why we weren't doing sequencing (samples were being tested for one thing and then being thrown in the garbage).
    - If you have public health at the federal level, and maybe through grants at state and community level, actually paying for sample collection and reporting, then that makes it sustainable. That is a good use of tax dollars.
- +1 on paying for data; likes another participant's phrase of trying to find data (strategic data stockpile) and what that looks like. There's a set of data that we should be thinking about (listed some of this before) and that participant made an excellent case for synthetic data as well.
    - We should learn from the UK government. They use Siren, REACT, and COG-UK (efforts to do epidemiologic characterizations that were real time that went above and beyond current data systems—getting excellent data about what's going on in the pandemic from them). We need to find ways to do active epidemiological investigation that would support key questions and understandings (primarily transmissibility, severity within key populations, and how the disease is moving) so that we can support ongoing understanding of how it's moving through our communities.
    - Need static data that you'll need regardless and then active data capture for the specifics of outbreaks (needs to be developed)

13. What are the lessons or successes from other models/approaches e.g. Models of Infectious Disease Agent Study (MIDAS) or Research and Policy for Infectious Disease Dynamics (RAPIDD)?
    - Malaria ATLAS project—maps of incidence
    - UK government (Siren, REACT, COG-UK)
    - From the standpoint of community engagement across The National Land Grant University Cooperative Extension System, they've conducted grassroots needs assessment for over 100 years and use that data to inform educational programs and evidence-based curriculum. The missing piece is how we're connecting to other health-related databases when they collect that information.
        - There's a national database of information that's grassroots led that is a success for them but has the ability to connect granular level data more broadly.

- One participant was personally involved in the creation of the MIDAS group for many years. The group initially accepted the mission of emergency support but then NIH decided that that wasn't their job (to provide operational support) and that there would be no explicit call to MIDAS investigators to provide decision support.
  - This points to incentive structure for participants
  - Everyone will try to help in an emergency, but it won't work if it's not aligned with the incentive structure (happened with MIDAS since they determined that they were primarily a research organization)
- The NIH RAPIDD example is somewhat similar. It was a shorter time duration, and it was targeted at supporting new innovative modeling. It was confined in what it was looking at. It was highly successful from an academic perspective but in terms of translating those academics into something that was operationally meaningful, that really wasn't the full remit of what it was trying to do. That's where the Center should focus: that translation of science into actionable capabilities.
  - How do you actually support responses and take the best from and support science, but translation to action is what's really needed
  - Learning from work done for MIDAS of improving the science, but moving forward with translating to action (missing component we need to work on)
- For the translational/T4 level (which is also going to lag behind T1 and R1), that's where we are right now. The National Extension System is in the process of translating science as part of an educational campaign around COVID vaccines. They are just now getting in place to educate at a different level (that's translation at the very grassroots level). There's a lag in time it takes to put new science into action.
  - Looking at clinical science alliance, there are four levels: T1 (bench research), T2 (translation to patients), T3 (translation to practice that includes training stakeholders and healthcare providers), T4 (translation at community level)
    - focus of levels is how to take bench research to the point that it gets extended to the general public
    - Note: this model was shared in the chat
- Some of this group's members come from DOD experience (60, 61, 62 kind of research all the way to 69). The fundamental idea is going from basic to applied research.
  - The Center should define who the customers are so that we can define what that translation needs to be (dictates what kind of results/evidence would need to be supported to discriminate certain policies/actions and then that dictates what type of modeling, needed data for collection, and what kind of systems need to be supported)
  - Also need to determine what are pain points of customers
  - There will be commonalities across all pandemics
    - We need to know transmissibility, severity, risk factors, and if something is really happening
    - Will need active epidemiological investigation capability as well
- T4: This is applied research where they study research and action and provide feedback on what really works in the field, what doesn't work, and how do we feed that back to the scientists so that they can continue to refine what works and what doesn't.
- One participant felt that after working at NIH for many years, they wanted to productize the results of their work (make it available and accessible to a much wider group). There's no good way to do that unless you go to the commercial sector and make proprietary software. Open source may not be the best decision.
  - Recommendation: Reconsider or at least define what is meant by everything open source.

- +1 for the comment above. Implicitly, people think that if it's open source then it's less expensive, but the software has implicit costs associated with it. For instance, if you think about Tableau vs Plotly, you need different skill levels for the different softwares. Plotly requires people to have internal skill sets.
- Who would be customers for this work?
  - Mayors of top 10 cities and governors of every state; authorities lie at locality level, so we need to support the people that have those authorities and have to be able to do that at scale
  - +1 for these comments.; Sees advantage of having multiple customers including creative ways to make this sustainable. Pharma may be interested in the data and may be willing to pay for it. Another example is the National Weather Service. Weather apps have real-time monitoring and forecasting systems, so it is a great example and gets thousands of people to engage and be part of their systems. For airlines, pilots also need a different visualization of that data.
    - examples: weather.com, Weather Underground, federal weather data
  - County governments and other key stakeholders need to be on board, depending on who they are listening to like local agricultural businesses, farmers, or local manufacturers. We have to think more broadly.
- How we collaborate with academia—trying to make all our data public in the spirit of transparency for how we work with the government at Google. To get to this point, we conducted extensive interviews, academics, policy makers. Decide what is useful and common denominator. Early access programs with data to academics. Get them to use the data to see what is useful. Revise and iterate and then make it public.
  - To a large extent, might be optimistic of what large companies could do. By and large we see ourselves as empowering others. The project program has been used in 1800 academic publications. Five authors are empowered. Would love to collaborate with a few academic groups. EG goal today to understand what data people would find useful to best address needs.
- One participant thinks having an open science orientation to this is really important—promotes reuse of data and learning from the group of modelers working and larger groups—seen with efforts of COVID forecast hub. Where people have used the data that has spawned other publications from people who want to understand the models better. Open data is really important.
- As an outsider from CDC/CSTE world—there are two camps from CDC (federal mindset) and STLT (local orientation)—people have different perspectives on the questions that are important. Having representatives from both of those spheres on the back and forth of the stakeholders, decision makers, and modelers to understand the targets of interest. Important to have both involved in that process.
  - Having an advisory committee with reps bringing all perspectives. Bringing on board a few state and local agencies and departments that could be deeply engaged to bring a proof of concept. Less examples of how state and local agencies work with 1-2 groups and fewer being supported by more groups. One or two collaborations with state and local groups to be used as a template to see what could be different.
- Needs to be a structure that supports interdisciplinary contributions in an ongoing way. Supportive of comments for needs for STLT representation. Many times, the models are put together at such a macro level that when it comes to decision-making within a state or local public health, they don't think the models are granular enough to be helpful. Public health is not aware of the different types of data that is out there to help support some of the decision-making. That's where having a regular interdisciplinary structure with regular collaboration and feedback will be important.
  - If we want to make it work, we need a couple of questions to focus on tangible items in addition to this broader structure to move things forward

        - Deeper dives in COVID-19 response, a lot we do not understand in modeling around people's behaviors about how they are responding to things. Need granular pieces of information like vaccines, vaccine up-take etc.
        - People say they are flooded with data and not using it
            - Certain types of data that we get a lot of but are not using it effectively and other types of data that we don't have on an ongoing basis
        - Google is happy to offer deeper dives on any of the data needed
    - When thinking about structural collaborations with workforce challenges we face, do you have examples of effective interdisciplinary collaboration to help skill development in service to answering important questions?
        - Maybe that there aren't great examples and that becomes a different challenge.
        - What do we mean by workforce challenges?
            - Observer: What I've heard, significant concern about the type of analytics expertise that is needed to make decisions about priorities and quick turnaround responses for current legislature needs, level of expertise. Issues being raised around future public health workforce, data savvy public health workforce.
                - When you look at a state health department, a small number of people on staff are looking at those issues. They are stretched thinly. It becomes a challenge to be able to respond effectively. HDs struggle to rapidly organize and process larger amounts of data to turn things around. Example: some of the data around the lab coming to HDs (trying to bring in health results and match to individuals with my data set to see how it changes the case). Those processes take so long, only being able to ingest and take in a portion of the data every day. The way the procedures to read it create backlogs to be able to ingest rapidly. Set up by folks not thinking about huge data feeds and how to bring them in really quickly. Sometimes it's the way the database is organized—not set up to run big reports and it tanks the database. Built some procedures that don't work well when you try to bring them to scale.
                - Patient matching in general is a huge issue. Not addressing in strategic and systematic ways. Less about forecasting (forecasting center won't have all that identifiable data)—speaks to issues of improving the data fields so centers can be more helpful. Timely data coming into public health needs to be looked at.
    - Ways to think about a program this Center could help support that trains a group of savvy data health analysts in public health. Support structure that is central, so they aren't abandoned. Could support modeling efforts in a local way. starting with Massachusetts Health to embed a student—think broadly to support a workforce like that.

14. Who are the stakeholders that are critical in the implementation of solutions?
    - The National Land Grant University Cooperative Extension System
    - Clinical Translational Science Alliances
    - Colleges of Public Health, Agriculture and Natural Resources, and Family and Consumer Sciences
    - Data providers—clinicians, diagnostic labs
    - Stakeholders: decision makers (health officers, mayors); pharma (when to make vaccine at scale)
    - Major Epidemiology and Public Health Organizations, e.g. APHA, ASTHO, CSTE, ASPPH
    - Academics, Public Health Officials, Elected Officials
    - If we just view this as a health issue, then we may not help the situation
    - Clinicians
    - Diagnostic Labs, Dx devices as close to or at the individual level as possible
    - Urgent care networks

- Cloud providers (Amazon, Google, etc.)
- State and local health officials
- Academic modeling groups
- Academic departments responsible for training relevant workforce (Public Health, Comp Sci, Statistics, Biostatistics, Epidemiology, etc.)
- Other federal government agencies who would be interested in the outputs of the models
- Federal health officials
- Private sector modeling groups
- Private sector data providers
- STLT, and those supplying data sets of interest—there often isn't good understanding of the metadata about the data sets themselves
- Private sector
- Liaisons—those developing public/private partnerships on both sides
- Policy makers
- STLT

15. What are other important gaps that need to be considered and addressed?
    - Agree with funding for data collection
    - Inclusion of community translational scientist and applied researchers
    - Consistent funding to support data collection, aggregation, analysis
    - Data use agreements, technology workforce in public health
    - Ability to do epi investigations that require collecting data such as genomic data, behavior, clinical outcomes
    - Effective communication of modeling results to inform public health action
    - Data scientists in public health
    - Rapid ability to work with the private sector in times of crisis—acquisition systems.
    - Research program on human behavioral responses to epidemics and disasters, and modeling of these behaviors
    - Coordination with parallel modeling centers across the federal government—how to build relationships with Economics, Labor, Education, Defense, Commerce, etc.
    - Funding for data collection
    - Leadership
    - Technical infrastructure for model coordination
    - Funding—how to support academic contributors
    - Prioritization of modeling targets/goals
    - Funding pathways—how to support public/private partnerships
    - Establishment of the framework and ongoing collaboration/feedback loop

BREAKOUT SESSION 2

16. What is needed to support and enhance the understanding and use of modeling, forecasting, and sophisticated analytics among the federal, STLT, and global public health workforce?
    - Facilitator: workforce needs and expectations of health department of epidemiologists that they may or may not have in this area
        - Probing question: How can modeling forecasting be made useful? What else is needed?
    - This resonates, it must be useful and easy to use for forecasters to do their jobs
    - This is a challenging question; must socialize use of quantitative modeling metrics (did this at Influenza Forecasting center)—this has been a slow and steady process where they keep a steady drumbeat of showing quantitative results, explaining them
        - This is uncomfortable; pushes boundaries but is really taken
        - Daniel Jernigan was quoted in the Washington Post offering a quantitative statement on flu season that sounded like a weather report.
            - Behind this one quote, there was six years of socialization
        - Cannot expect that this use will take off overnight, but does require people in CDC to slowly and carefully socialize the use of these models and explaining what the outputs mean
        - Also incorporate inputs and explanation into the reports presented by CDC
    - Would be wonderful to have forecast at zip code level and not just at the national level

17. What is working well to support the use of modeling and forecasting for public health? Are there things that aren't working that should be addressed or avoided?
    - Pooled testing in schools that's happening around the country is new and working
        - This is giving parents confidence
        - Would be a shame to lose these advances once we are out of the pandemic phase
    - In the COVID forecast hub project, what's worked well is having a coordinated effort where the modeling objective is clearly defined and where multiple groups can be working
        - Specific objectives and many teams working synchronously
        - This allows a model coordination center to pool perspectives on multiple models and synthesize them and present agreements and disagreements
        - Diversity of perspectives from different models and groups is important
        - Having the same objective is really critical in this process
        - Pitfall—Having 35 to 40 groups answering the same question—that might be too much; good to diversify questions and efforts. Nobody is left to answer other interesting questions. Avoid sucking oxygen out of the room and leaving space for both coordinated and autonomous efforts
        - Having different and spread-out efforts would be great
    - What is not working is that results of diagnostic testing aren't being reported.
        - If we could take the work on flu or COVID and expand it to other respiratory diseases, that is the future

18. What is needed to support training for decision makers and frontline staff to improve public health decision-making?
    - One participant noted that he did not understand the question
    - Model results need to be shown to them regularly and often, e.g. during weekly updates. This helps with socialization of these kinds of quantitative outputs without requiring special training.
    - Workshops that bring together public health practitioners and modelers to hear the different perspectives

- Model evaluations need to be shown to decision makers to help them build trust and intuition about them
- Occasional workshops that could bring together decision makers to talk about how they have used model outputs in decision
- Teach modelers how to communicate to decision makers
- Frontline staff should get useful feedback on data provided
- Make it user friendly with minimal requirement for training support
- Use community translational scientists/practitioners to work with modelers on communication and messaging
- Funding
- Research-based training curriculum, consistently used
- +1 for the comment on making training user friendly
  - Showing results to people and having it be a part of the process is crucial—make it part of the daily cadence
- Decision makers don't have time for training. Should be part of daily communication with decision makers.
- Put these into trainings for frontline workers, etc. Has access and there is an effective interface with modelers and CDC frontline/decision makers to foster collaboration. Has participated in workshops with CDC, modelers, and CSTE.
- It is an art to communicate to decision makers. Training to modelers specifically would be helpful so they can communicate to lay audience.
- On decision-making side: here's what the model is telling you and here are your options for making a decision
  - These need to be presented at the same time so that there are useful and meaningful options as well
  - On the public health side, the recurring question is "when will this be over?"
  - We need to develop meaningful responses from the model

19. What are the lessons or successes from other models/approaches, e.g. Models of Infectious Disease Agent Study (MIDAS) or Research and Policy for Infectious Disease Dynamics (RAPIDD)?
    - Coordinated models/groups/perspectives trying to answer the question. Wisdom of crowds works well; nobody working in silo.
    - National weather service is a great model for ID tracking and forecasting
    - Include diverse stakeholders
    - The new National Wastewater Surveillance System is a great model that should be sustained
    - Credit card fraud detection. Amazing example of finding signal/noise in real time.
    - The Weather Underground app is a great model—maybe we could create "the viral underground" network of citizen reporters?
    - And dedicated funding for key data modelers needed. CDC should be paying for data.
    - Success: a wisdom of crowds approach is valuable for modeling efforts
    - Dedicated funding efforts for modeling help build community and drive innovation around these questions
    - Agreement with the need for funding
    - Leadership for coordinating communication
    - Regular, ongoing communication between the modelers and public health so they can be continuously tweaked and revised

- For communication to the public, good basic graphics for the public become critical. Public is used to basic kinds of visuals; this contributes to the success of something that is meaningful to them from the model.
- The models are only as good as the data coming in—much more investment needs to be made to improving the data available to public health especially the timeliness and completeness of the data
- Are we talking about specific communications lessons? Yes.
  - There were some key takeaways in the first discussion related to this: both efforts (MIDAS and RAPIDD) had model coordination efforts
  - Model coordination is really important—both of these efforts had models coordinated to ask the same question and that's a success of these groups
- Funding has emerged as a theme on the Jamboard
- Public is habituated to making decisions off basic visuals—important to success of understanding what is meaningful to them in the model
- Underscore the comment about how models are only as good as the data coming in and CDC must pay for that data

20. What is needed to improve the communication of forecasting and modeling results and information to decision makers and the public?

- What we have been doing—producing local illness score based on hyperlocal data, success with engaging consumers because they are able to see impact on community and how to adjust behaviors
- Need to be able to make data communication/forecasting relevant to their communities, able to monitor illnesses in their communities, obtain services. Hyperlocal relevance and actionability for communities.
- Communicating consistent, shared messaging. Eliminating academic/research jargon, increasing accessibility.
- How to translate models into something meaningful (e.g. what if the model represented something in the past that has held meaningful information for communities?)
- Understanding what we are communicating is not just a health problem, underlying the importance of economics/social sciences that influence human behavior. Creating relationships between subject matter experts in different domains, communicating infectious disease dimensions and other sectors. Worthwhile to communicate simulations to fed government, how will this Center interact with them
- Pulling in trusted communicators. eg. communicating insights so that a trusted member of the community can relay. Question is, how do we make sure they have the right real-time info?
- Facilitator clarifying question: Is there a difference among audiences between forecasting and modeling?
  - Packaging—it's about creating the right relationships in different spaces and how they are able to receive the data (usable insights for experts and for the public)
  - Packaging with suggested actions for the public, whereas for decision makers, they would need to include projected impact from actions, etc.
- Circling back to trusted sources conversation—campaigns should include identifying trusted sources, people with influence. Message needs to be carried outside of the healthcare side (e.g. working with faith-based leaders, people within farming communities, workplaces, etc.)
- Empowering sources of health-based knowledge, health-adjacent fields

- Distinction between counter-/factual forecasting—they can get muddled in communication, needs to be clear what the audience is seeing. What statements of the future are we making, and how to evaluate (e.g. scenarios where mask use and school closures change. This changed over the pandemic)?
- It's lost in communications efforts what the forecast horizon is and what accuracy they reflect (e.g. 1 to 4 weeks versus 6 months when several months into the pandemic)—expectations around accuracy and understanding/interpretation. People tend to believe forecasts may not always work or have a magical element to them.

21. What is needed to increase two-way communication between modelers and decision makers to better define the highest priority questions for modelers to address?
    - Expanding understanding of existing platforms and tech to not reinvent the wheel—integrating different existing platforms
    - Different kinds of decision makers have ways of optimizing their own decisions. Engaging client as early and intimately as possible—see people as part of the team.
    - Unclear to forecasting folks whether use cases are aligned to how they're measuring accuracy (e.g. is a forecaster interested in upticks, periods of inactivity, etc.?). Do you trust a forecast to be active during an upswing? Is it more useful to have accurate forecasting during a lull? May be helpful to see these as different use cases in order to see impact and value for forecasters.
    - With clients we do monthly consultations, or as needed, to prepare to dive into the data. Translating insights is a big point. Our data is the client's data (think account management model), and it helps with translating the data and making them usable, having a partner in that understanding.

22. How should these data (forecasting and modeling results) be shared and are new tools/ platforms needed to facilitate?
    - Take stock of existing platforms (public/private sector and academia) to avoid duplication
    - Real-time integrations of signals/models into public health dashboards (perhaps via API)
    - Granular! Granular! This is critical. The best way is probably by creating synthetic data sets from the individual data.
        - Can be both—Kinsa aggregates hyperlocal/geospatial data from individuals (aggregated/ anonymized for privacy)
    - Consumer engagement focusing on takeaways—how does this impact people/are the i nsights actionable?
    - Facilitator clarifying question: are "granular" and "real-time" related?
        - Not necessarily, though for spatial and individual-level data you'd like it to be real time.
        - For us it's both since we'd like to be able to use it for our hyperlocal analysis. So, they can be the same.

23. Who are the stakeholders that are critical in the implementation of solutions?
    - Private sector/innovative data sources
    - Modeling leads at agencies other than CDC
    - Including biodefense space/DOD
    - Public health officials and political leaders—what is the threshold for action? Can we tailor models to clearly denote whether we meet this threshold, which is often political?
    - Elected officials at the state, county, and municipal level who have decision-making authority
    - Major public health professional groups APHA, ASTHO, NACCHO, CSTE, ASPPH, etc.
    - Education organizations—critical for influencing behavior

24. What are the bold, meaningful, and feasible implementation and utilization goals to be accomplished in the next 5 years that were not previously mentioned? (refer to responses to large group discussion Q2)

    a. Which are the greatest priorities and why?

Research Gaps

COLUMN LEFT BLANK

Data Gaps (that can be filled)

- In addition to the variety of data, geographic scope matters (ranges from as local as possible and insight into global data that may offer guidance into the new threats)—applies to all data we are trying to obtain
- Metagenomic data
- Granular data collected at the state level (data given to CDC is usually less granular)—synthetic data sets generated on state data should be made freely available; avoid tracking individuals and respecting privacy; work around data use agreements
- Would others agree on the value of a large table that, by county/age/possibly race/ethnicity includes the numbers of persons with various comorbidities and their overlaps (combos of 2, 3, or more)? This was a big challenge in vaccine planning for both 2009 H1N1 and COVID-19.
- Race/ethnicity and other intersectional data is missing nationally
- Urgent need for us to bring together better public health data that includes more information about the populations, including vulnerabilities, social services, etc.; put this together with healthcare data that can be linked—will lead to effective analysis and forecasting
- Obtaining granular data is important; we should also gather data on the structural drivers of health inequities and health outcomes—collect information on multiple indicators to support work on SDOH

Implementation Factors (i.e. policy aspects, system, structural, workforce, etc.)

- National syndromic surveillance platform as a bold goal
- Legal reporting requirements for social determinants data—when there isn't a legal requirement, that causes data gaps; during COVID, there have been challenges and concerns around providing data for certain populations that's not available; labs aren't set up to even collect this information (infrastructure is absent because there is no requirement to report or funding); long-term goal, need national or higher level requirements on reporting (varies state-by-state); levels of detail do fall off the more the data is aggregated
- The public health system needs more and more robust informatics workforce.
- Support for data coming from health care—claims and EMR data can be extremely valuable and timely sources of information and linked and combined with other important features. Not only serving as early indicators of activity, but alternative measures of activity and alternative measures of ground truth entirely
- Biggest implementation factor in academia is that incentives of doing this kind of work (building real-time models meant to be operationally tracking real-time infectious disease) don't exist; if we are relying on this work to be done in academia, challenges will arise in that academics will work toward publishing; the type of work that's most effective isn't always work that should be turned into paper—often getting into weeds and details of problem doesn't fall into traditional researching and modeling (doesn't mean this work shouldn't be done—just needs to be considered when discussing incentives)
- Another participant added that the same can come from private sector—novel data sources can increase timeliness of forecasting and reporting (e.g. Kinsa is communicating with underserved communities and sentinel groups very early in the illness cycle)

- Policies around data provenance and usage (especially in the context of big mixed data sets from various sources/data lakes)
- How is health data captured in rural, migrant, and underserved populations at a very basic level?
- I think public health departments need to build up workforce—for many jurisdictions it falls under the general "epi" groups who are not always able or trained to do this work. Dedicated funding for modelers would be best.
- A critical aspect of developing policy based on models is how community partners can be involved in educating the general public as part of a continuous loop
- Supporting training programs that merge public health data challenges with advanced statistical modeling and analytical techniques is critical to help the field grow

Other comments:

- It's probably bold, not sure if feasible—national patient identifier
- Outbreak analytical support for every public health officer in the country.
  - Seeing how King County (Washington) used Institute for Disease Modeling to help them think through a range of different questions for Seattle and Washington State developed a really good relationship in developing the different analytics to help them think through the different policies that needed to be put in place at the city, county, and state level. We should think about how to actually scale that capability to other cities, counties, and states (should be a focus of the Center).
    - We should also learn the lessons from what that interaction was like and try to avoid things that didn't work well and amplify things that did work well
    - We need to have a much stronger, faster capability to bring in private sector capabilities into public health and government. Government acquisitions move much too slowly in times of crisis.
    - A part of that has to be an acquisition problem (and how to do acquisitions differently) and another part is funding to make that consistent
    - Bringing the two together in a more robust way is needed
- To build on what DG said, maybe we should establish channels of communication between the government and public sector. We need mechanisms to reach out to stakeholders, especially in times of crisis.
- Based on what another participant just shared—wondering if there are thoughts regarding national standards, systems that allow for and provide for consistency?
- In Seattle, they've been fortunate to have lots of huge players that are close in proximity that have vested interest in them. A lot of this came through emergency response and opened up opportunities while building skill sets (like working with a forecaster from Starbucks on modeling). Some are companies that he never thought he'd be working with.
  - It's hard to invite new people when you are stressed. Sometimes it felt like drowning with all the support groups because they don't have good infrastructure in place to track those that have reached out during the crisis.
    - Amazon, Facebook, Starbucks, and IBM all reached out
  - It would have helped if things were set up beforehand
  - Need a clear understanding of roles and needs
  - Data sharing process took a long time with IBM, so it can't be done with every organization
  - Each department and each jurisdictional entity should acknowledge if they have key people in place that their sole job is public/private relationships specifically to data. We need actual data governance people that are embedded in public health departments.
  - There is a lack of data transparency standards nationally. We've been relying on NYTimes and Johns Hopkins as resources when public health should be able to take that on.

    - Basic descriptive demographics are a must
- +1 for the above comment

25. What structure or design would best achieve the desired scope including the coordination and collaboration with state, territorial, local, and tribal (STLT) public health entities, academic, health care, private industry, and interagency communities?
    - Need data sharing agreements and public/private relationships specific to data
    - Would like follow-up on what it would look like to have some type of centralized resource or infrastructure that would provide consistency and governance
        - What are the national standards? There are different levels of capabilities throughout the country. Large companies that are not disease-focused may have skill sets that could help. They are partnering with the county because they have control (decentralize government). There are also special interest projects that they do with companies that are local, but there is a need for some sort of national position/national guidance. Data sharing is actually a roadblock because they have to curate everything that they want in a specific way and may actually lose a lot of the information and capabilities along the way. Positions they have are much more political. Some may be able to do that on their own, but that's not feasible.
            - Another participant agreed that it takes a long time to put the paperwork in place for data sharing, which is why all of their recent data releases were just made public
        - To build on that, it was challenging with a reporting mandate to get good, quality, clean data for situational awareness at the national level just because of the question of authority and to be able to capture these data in a meaningful way even in public health emergencies
            - For the majority of the funding that goes to state/local public health agencies, a lot of their funding agreements don't allow some of the research component of data (this may be changing)
            - It's a good reminder to think about what activities are allowed with funding for public health agencies
    - Governance to support information flow:
        - You can come up with standardized ways of doing data use agreements. You can do a lot of the heavy lifting ahead of time.
            - Most important thing is that you need to (instead of coming up with a governance structure that would force counties to participate and provide data) get to the positive feedback loop (virtuous cycle of providing them something that will help them to do their job better, faster, and stronger). Then they will want to help and provide the data and you're giving them something that is useful so it's a win-win situation. It can't be seen as an academic exercise; it has to be rooted in helping the response move faster. If we can get there and find out what the critical use cases are, then we're good.
                - Should be NSF or NIH
    - A lot of places around the country do not have strong informatic support. There isn't a lot of funding or expectations of either state or local jurisdictions to be able to really take this on. So, the burden gets added to others. Typically, leaders would be physicians and have a couple of data folks, expectation to do core informatic projects is placed with the same people that have to do public reporting, investigate diseases, do some additional research, analyze what's happening, etc.
        - There needs to be a lot more money and informatic support at every level of public health agency
        - There is a shift, but we're well behind for basic informatics needs
        - National person index: many jurisdictions have this for their county, but it requires years of work and needs dedicated staff that do this every day because things change all the time (but a lot just do this as a side project)

- +1 with the above comment; also one of the main concerns is to make sure there's clear guidance and potential governance and downward policy/support for changes at the state level because funding that is supposed to go to informatics, does not go there
  - Define what public health informatics is
  - Funding/capacity and capability building is necessary; needs to be a unit within agencies
  - We are in jeopardy of repeating this again, but we have a lot of opportunities right now in the data modernization initiative
  - Looking at the centralized data source is important, but we have to take into account that anyone that has received funding to do the data modernization work has competing priorities, so now is the time to identify key data that will need to be captured or considered in the systems. Without the expertise at those organizations to select the right systems, articulating requirements will be challenging. If we don't have expertise at a certain agency, then we need to find a way to provide assistance and technical support to make sure that they're ready for the next phase.
  - Agreement with this comment expressed by participant

26. Who are the stakeholders that are critical in the implementation of solutions?
    - Public health services groups within each state (public health laboratory, local health department, state health agency) and groups such as organizations that are responsible for managing a centralized information technology services group
      - Information technology vendors, academic institutions, big tech and data firms, CDC, other federal partners
      - Make sure there's coordination across federal entities about what needs to be collected as well as validating the ability to collect those data and determine within what system do you collect those data
    - Clinical providers

27. What are the important implementation or utilization gaps that need to be considered and addressed?
    - Difficulty in getting large-scale data about health outcomes at the level of zip codes/counties/census tracts
    - Data use agreements
    - Nimble acquisition systems, or said differently, need to create a sustainable market for private sector to stay involved in support public health informatics
    - Providing consistent usable modeling results that will help health officers effectively manage a public health emergency; have to make their life better somehow or will not be worth it
    - Technical, data science workforce in public health
    - Effective communication or translation of analytical results to guide decision makers
    - Older systems that are not readily adaptable
    - National standards on core/minimal data elements
    - Staffing/partnership identification
    - Assess and take advantage of existing technology/investments—doesn't always have to be a new solution—use what is working

28. What kind of staff are needed in the implementation of solutions?
    - Policy, legal, ethics, expertise that can advise on policy levelers and aspects
    - Data scientists, statisticians, machine learning experts, software engineers
    - Technologists
    - Data scientists

- Someone who loves data use agreements
- Super swell people that are just good to be around!
- Informaticians at every level
- Communications/PIO
- Public health data subject matter experts (in general)
- Lawyers

29. Who are the additional stakeholders that are critical in the implementation of solutions?
    - Range of end users at community level—outside public health that rely on timely, accurate data
    - Mayors, governors, public health officials
    - Big tech [agreement voiced from another participant]
    - Community groups/activist groups
    - Big chain pharmacies
    - National professional membership organizations (AMA, ACLA, APHL, CSTE, etc.)

SUPPLEMENTARY COMMENTS (SUBMITTED AFTER SESSION WAS COMPLETED):

- Participant 1 on email: I'm not sure if this is buried within the notes so apologies if already discussed. As I thought about how federal programs could modernize the data infrastructure, I wondered what this analogy holds. There were significant incentives in place for adoption of EMR technology several years ago. If this could be done at a minimum to fund standardized collection into state, county, and metro public health departments it would be a great start. This information could then flow into a different kind of Health Information Exchange (HIE) to use another analogy at the state level. If the infectious disease informant exchange were done in an anonymized way, then testing organizations, enterprises, etc., would also be able to contribute.
- Participant 2 on same email: As Participant 1 kindly noted, our research group at CMU is working on the importance of collecting and integrating infrastructure data (both physical and digital) to inform pandemic forecasting and modeling.
    - I have attached a brief one-pager making the case for more data on the built environment in informed more rapid and targeted pandemic response (with references if needed on page 2)
    - Happy to follow-up if it would be helpful and/or upload this comment elsewhere if that would also be advisable in response

To enhance the flow of the discussion, questions 1 and 2 from Listening Session 1 were swapped in Listening Session 2. Question 2 from Listening Session 1 was also slightly modified so that it clearly demonstrated a link to Dr. Jernigan's presentation and framed the discussion in the context of the proposed Center for Forecasting. Question 3 from Listening Session 1 remained the same in Listening Session 2, but a probing sub-question was included to inquire about lessons learned between infectious and noninfectious disease modeling.

DMI LISTENING SESSION 2
WEDNESDAY, JULY 28, 2021

MAIN SESSION

1. In thinking about the future of forecasting and modeling for public health and the Center for Forecasting that Dr. Jernigan described, what are bold, meaningful, and feasible goals to see accomplished in the next 5 years in the following categories?*

   **Q2 from the first DMI Session, slightly modified. Moved to be the first question in the second DMI Session.*

   Research:

   - Use the experience from the COVID pandemic to evaluate the success of different approaches, models, forecasts to actually predict what happens
   - We don't know what the next threat is going to be, so it's important there's a broad attempt to model and understand the endemic and pandemic infections, not just focus on the pandemic. COVID has demonstrated the uncertainties.
     - Data gaps are large—manifest; we need routine surveillance (pathogen and serological) in the US and globally
     - Genetic surveillance (of pathogens) is also needed and needs to be widely set up
     - We need strong investment
     - Success of modeling depends on provision and openness of data
   - Data standards across the board; being able to exchange data bi-directionally via automation/interfaces vs. manual uploads/downloads
   - In addition to revamping public health reporting, it's important to supplement it with found data—data not created specifically for reporting (i.e. EMR, medical claims)
     - The advantage of this data is that it is broad
     - You don't have to set up specific systems, they are quite broad and can be queried at anytime
     - Would like to see system that collects EMR data throughout the country—fine to leave line level data at its local jurisdiction as long as it can be queried—don't have it right now and could be a huge contribution
   - Agreement with comments about data standards; we should make sure investment is broad because we don't know when the next threat is
     - I would also recommend a campaign to focus on accurate forecasting of SARS COV2 (e.g. talking about COVID-19)—it will be with us and will be a high priority for awhile
     - There's a high demand for these forecasts—part of what we've seen over the years while trying to forecast infectious diseases is that we've [e.g. modelers, forecasters, etc.] been interested in doing it but no one has been interested in seeing it
     - [Now people are interested in seeing these forecasts] which reflects a change; we need to go after accurate measures on different time and spatial scales and that will push the field forward

    - We need to invest in good data—in UK there is good infection data rather than case data; clarifies debate because it's easier to discuss something that comes from a simple sample
    - In terms of the actual amount of money you need to fund good infection data, it's a lot more than the money you need for modeling studies, but regardless it's important to move everyone forward at the same time
        - Agreement expressed for this comment
- Ensure research capacity includes modeling at several levels, from ecological/population-level prediction to individual-based models (like ABM) that incorporate variability in individual human actions and interactions that influence emergent social phenomena
    - Ensure we have capacity and expertise for types of modeling, especially at the population level
    - Need individual-based models and network-based models that show how individual interactions influence the whole
    - Demonstrate how emergent social phenomena impact the wider population but also to explore how disease spreads based on individual actions, choices, policies, etc.
    - Must ensure different types of modeling capacity
- One thread that another participant mentioned is the importance of thinking deeply about serological surveillance as cases fall, we could be in a good situation where there is very little susceptibility or a very bad one. Would be entirely invisible unless you have data on sero surveillance. Found data side could be exploiting existing threads of blood banks through to the cattle design that another participant echoed.

Data Gaps:

- Regarding data gaps in the US, it would be very helpful to understand what types of data are anticipated to be needed so that people who are conducting surveillance are more familiar with surveillance and the provision of data needed for forecasting
    - Line level data or not line level data?
    - Must address barriers to data sharing (legal, etc.)
- We need a very substantial investment in state and local public health in the US—we simply do not have the capacity to generate the needed data currently.
    - This is under discussion—but need people to help build a robust data infrastructure, run, and maintain systems, analyze data, and collaborate with modelers/forecasters
- Social determinants of health with race/ethnicity, household crowding, and underlying medical conditions

Implementation Factors (i.e. policy aspects, system, structural, workforce, etc.):

- How big a workforce is needed? How much work is done on this?
    - There is a big workforce gap if we want to work on this; presents a challenge
    - You could work out a back of the envelope calculation for the workforce needed for this relevant to the current challenges and based on current capacity
    - Britain is ahead in this, but are there shortages in the UK
- There is a challenge in the UK around government capacities
    - The UK is undergoing a large increase right now if you think about the new government groups doing this
    - Massive expansion is underway
- Participant is in academia (framing)—we need to grow sectors in many ways, not just in current disease modeling groups but we need to grow all workforces across universities and other groups
- Plug in for not just needing modelers but also those at the interface of modeling, decision-making and policy makers

    - These people are go-betweens (good communicators) who ask important questions about what modeling can and cannot do and help with the interpretation
    - Would be able to point out where models have gone wrong in terms of generating forecasts
        - Support expressed by another participant
- Participant strongly agrees with above comment
    - You need translators and must have dialogue in public health at various levels (local and state) to feed questions in and interpret and understand results
    - You also need people to work with elected leaders and decision makers to contextualize the results
- Arrangements we have to do with data agreements, in our state and outside our state, and statutes with employees, etc., needs to be reviewed and reassessed
    - These regulations are outdated and limit researchers and others who work with data at multiple levels (and hinder the ability to share)
    - States had to put up executive orders to get the data they needed to respond to the pandemic appropriately—particularly challenging in the US
- It's a general argument about secondary analysis—we have many current opportunities to expand this data and investment is relatively cheap—should be a priority
    - Sometimes overlooked that exploratory analysis—primary data is more expensive to procure
    - We must get faster consensus on a key point (particular to infection, etc.) and investing in primary data assists with this
    - We need to keep an eye on how budgets have changed and ways to get new data that might be needed
    - In UK, currently working on random sampling in the community around respiratory infections through serological surveys (SR's experience)
        - Support expressed for this comment
        - There are other ways to examine social behavior, etc.
- One example of new data needed: one of the most important things they need to understand is the relationship between vaccines and natural immunity against COVID and onward transmission. Hard thing to get at and not easy or inexpensive to do. There are ways to do it with clever cohort studies linked to modeling.
- Agreement with the comment that we need non-modelers who can translate and inform models. And, regarding workforce, need education of the current and future public health workforce (e.g. at state health departments) about the value of modeling and forecasting. Many working in public health do not see the value of this work because they have not received formal training on modeling methods, impact, and limitations.
    - Around education for workforce—the workforce that is currently present for state health departments—we often hear frustrations with modeling and forecasting as a whole
        - Need to incorporate into curricula for public health practice and inform current public health workforce members what modeling is, what it impacts, and what are its limitations
        - We shouldn't reject modeling as a whole because of its limitations and assumptions—there's a way that can be tied into this
- One other general point is we need to expand and increase our health information exchanges and increase these partnerships between HIEs and public health entities across the board
    - Support expressed for this comment
- Need to improve interoperability of public health to line level health data
    - Would address issues of data lag
    - Line level data on hospitalization, death, etc., could be improved in real time (in terms of availability, etc.)

- Interoperability is needed in both the short and long term
- Electronic case reports exist, not always in place
  - Support for this comment. These healthcare data are massively valuable assets. I think you need functions in the new organization focusing on how to release those data.
- Agreement expressed. There is significant amount of very valuable data that's hard to get your hands on sometimes

2. Related to forecasting and modeling, what are essential or priority public health questions for analysis in the shorter term (1 to 3 years)? Includes sub-question a.)*

**Q1 from the first DMI Listening Session.*

- e.g. impact of COVID-19 NPIs on dynamics of other infections
  - Evolutionary dynamics of COV2 (hard question)
- Key questions are what would be the impact of following certain public health policies or not following them
  - What we are running into on the ground is a large portion of the population has now signed off COVID and are finished
  - If we are thinking of COVID in particular, what is the impact if we can't turn people around and get them to be compliant
  - In future outbreaks and pandemics, we will have a similar challenge
- Question is more behavioral—what was/is the impact of modeling on decisions that are actually made? What are the views of elected leaders on modeling?
  - There's a gap in understanding and evaluating the validity of findings and predictions and acting on them
  - Important topic—if modeling is to be useful, it must be understood and used in decision-making
- Refinement of comments already made—relative transmission of different places, such as in the UK
  - There are seasonal and strain changes, which affects the up and down movement of disease
  - Potential for fruitful work to examine relative transmission changes—wide scope
    - persistent high transmission areas that need to be understood—there will be specific findings in this area that can be carried forward

a. What questions for analysis might be needed over a longer time, over 3 years?

- Useful to break down analysis into different stages
  - Step one: True surveillance (surveillance of unknown—which is different than what comes later)—November 2019
  - Need novel data streams, be truly global, estimate the probability of different outbreaks happening in different places to allocate resources appropriately
  - What modeling and forecasting is needed when a local outbreak is detected?
    - Second step: February 2020—not clear it's going to be global pandemic but there is concern. Elevate initial findings and examine concern.
  - Third: epidemic and pandemic and just tracking quantities in different places
  - Math, questions, and assessments needed are very different in these three phases and should be considered differently
- Map between genotype and dynamics. Clearly something that is challenging. De-bridge contrasts between MERS and SARS.

- Exploration of the changing interaction between human and animal populations in natural and built environments and its potential consequences for introduction of novel pathogens into the human (and animal) population
  - Explore root causes of pathogens moving into animal populations
  - Example is movement of pathogens from animal populations
  - How might we change risk behaviors due to potential of spillover
- Human behavior is huge factor here in success of huge pathogen to spread
  - Is there a way to model human behavior/willingness to adopt certain practices/restrictions depending on public communication methods or messaging?
  - Can we model human willingness to adopt certain practices due to willingness?
- Underestimating probability of outbreaks: suggest explicitly including recent/current airline/mobility data. These data have been difficult for us to access at the local level.
  - Mobility travel data—available at the national level but not at the local level, which hinders us (at local level)
- Agreement expressed with the above comment
  - Data that are modeled need to be made available at local, state, and federal levels
  - Important to generate and use data at local or community levels
- Short and longer term: infectious disease interactions with weather, climate, climate extremes, etc.
  - Human demography, human dispersal—these are long-term issues that should be modeled but are feasible to model if we have the right data
- CDC has large global health surveillance programs that are intended to build capacity and have ample funding both domestically and abroad
  - Building out EMRs, surveillance—despite some limitations it would be good to have modeling integrated into systems so that they can be utilized
  - These grants lack money in peacetime efforts—need to build structure during these times so that tools can be used during concern periods to address public health and clinical response
  - If a concern appears, then money will flood in

3. How should infectious and noninfectious disease modeling research and application be prioritized and enhanced? Includes sub-question a.)*

**Sub-question a.) was added after the first DMI Listening Session*

- Comment on wording—when we say "infectious disease modeling," we are referring to mechanistic models though there are many other models that should be considered as part of infectious disease modeling that have been left out
  - There is also a point to be made about noninfectious disease/epi
  - There's synergy if we look at and question data sets
  - Recommend doing this before drawing a line between infectious and noninfectious diseases
    - Two participants expressed their support for this comment
- Priority is data improvement—must go hand in hand with refinements to modeling
- Agreement expressed for first comment; there are many tools (trad, epi, biostat) that can be brought to bear on a range of health problems from a non-mechanistic standpoint
  - Dynamic systems—fundamental component is that it is communicable and contagion process
  - there are other systems that have been put into the contagion box (such as with opioid crisis)—this could open opportunities for dynamic modeling approaches to bear on system
  - Use mathematical modeling approaches on this side of things

a. Are there lessons from one that can be applied to the other?

- Social behavior or health-seeking behavior
- Agreement expressed with the above comment; engaged on project with trying to model and forecast suicide like a contagion
  - Challenges: data collection, data availability, what's available in a set of outcomes that takes place
  - Highly analogous to one another
  - Lessons that can be applied: data is critical, so if you don't have observations of phenomenon you can't disentangle what's going on, simulations are going to be less accurate, forecasts similarly less accurate
  - If you move into space of noncommunicable diseases, there's abundant data, however in other instances it's very difficult to access and negotiating must happen
  - Once you have data, another set of questions arises about the helpfulness of the data and the modeling that can be used
    - Agreement expressed for this comment
  - Building on earlier comments, this participant would claim that access to data is currently the biggest bottleneck to progress, more so than methodology, algorithms, or models. They are curious whether others in this forum agree.
  - Issue was raised with sero studies; but even in noninfectious diseases, populations that aren't being diagnosed are due to a variety of challenges (such as the appropriateness of testing that also warrants examination)
  - What we haven't mentioned is impact of disparities on health equity
  - How does the prevalence of underlying conditions in populations, access to care, poverty, zip code affect infectious and noninfectious diseases? *(also a data gap)*
  - At the local level, the community level—what are the conditions that exist in the population that result in disparate outcomes?
    - Support expressed for the comment above; these issues need to be better integrated in modeling and forecasting
- There could be more discussion of ways in which HDs, hospital associations, and other regional systems share data
  - Create a data repository so that we can build on what was learned during COVID and continue this work with these organizations
  - Have these relationships and place and create access to data that wouldn't normally be shared

4. What new or innovative data sources should be considered?

- Better pharmacy data
- EMR via Health Information Exchanges
- Social determinants of health data
- As another participant said, better travel data, especially international travel level, at granular geographies
- Technology for mass random collection of pathogen and serological data from individuals: huge power for inference, detection of asymptomatic infections etc., though key to get data security right
  - Support expressed for this comment
- Better data on mortality and morbidity by sex, race, and ethnicity
- Would love some routine sites of active surveillance (as opposed to standard, passive surveillance)

- Random serological samples tested using assays that give stratified immune levels
  - Support expressed for this comment
- Disaggregated race/ethnicity data—the 5 or 6 main categories can be inadequate
- Sentinel surveillance systems, e.g. testing everyone who shows up for health care (could be anonymous)
- All routine surveillance data open and refreshed every day
  - Support expressed for this comment
- Timely reporting of outcomes
  - In research, built out system of forecasting for West Nile—did with accuracy but real-time data had 1- to 2-week lag from collection to actual reporting
  - Case data in humans had 3-month delay in cases being reported which impacts functioning
  - Enhancements of pipelines in clinical infrastructure that enables faster reporting
- These data collections—much more powerful if global
- Basic risk factor data (e.g. healthcare personnel, long-term care yes/no) collected on laboratory submission form for reportable conditions and reported to health departments. Through LIMS.
- Timely data on use of protective measures
  - e.g. masks, social distancing
- Immunization data
- Metadata—requests to provide all surveillance data everyday with no barriers
  - Surveillance data barriers are sometimes logical, but we should examine guardrails and caveats and how we should be able to access data (missing from this discussion)
- Global data for vaccine preventable diseases
  - Tradeoff between immediacy and timeliness and quality; support expressed for the above comment
  - Timeliness is much more important than cleanliness and accuracy
  - Can correct for noise and bias of early data
  - Make case for worrying more about getting data out quickly—corrections can come later
  - Build infrastructure and modeling system for being aware of early noise and bias
  - Agreement expressed with the above comment; not just earliness but corrections can apply retrospectively to large swaths of older data
    - Must provide metadata about timing and nature of updates
  - Yes, providing extra metadata might be solution—must also have process to understand how data are generated
    - Understanding how this works at CDC and local levels is important
    - Transparency and more detail of how data comes to be and what it represents and how it represents in modeling is critical
- Live, aggregated, representative healthcare episode data
- More data to assess the health and economic impact of infection and noninfectious diseases
- Support expressed for the above comments
  - Biases in data that should be included in metadata and must be communicated to modelers
  - Biases and timeliness need to be addressed
  - In general, in the US, we have a very weak connection between healthcare system, healthcare data, and public health data
    - Public health data often go to modelers
  - Need to improve interoperability (copied response from above) of public health to line level health data

        - Would address issues of data lag
        - Line level data on hospitalization, death, etc., could be improved in real time (in terms of availability, etc.)
        - Interoperability is needed in both the short and long term
        - Electronic case reports exist, not always in place
            - Yes, these healthcare data are massive valuable assets. I think you need functions in the new organization focusing on how to release those data.
        - Support expressed; significant amount of very valuable data and hard to get your hands on sometimes
    - Also need standardization (and centralization or organization) of data
        - Agreement expressed with this comment

5. What is most important in the identification and prioritization of research and infrastructure development?
    - Noteworthy article: https://pubmed.ncbi.nlm.nih.gov/34255055/—Aggregating Electronic Health Record Data for COVID-19 Research-Caveat Emptor
        - Call to aggregate EMR data—this article examines challenges
        - Plea for preserving data quality and interpretation at local level as we try to rush and aggregate data at large scale
    - We've had a huge amount of discussion about data, which is important, but what is really needed is integrated systems approaches where we are building out analytic practice, modeling, and forecasting simultaneously
        - Agreement expressed with this comment
    - We need things like complete data through laboratory reporting
        - Data through laboratory reports are often missing important pieces of data
        - Reporting from healthcare sector needs to also be considered
        - Fundamental challenges—implement interoperable and automated systems across the board
        - Vital statistics haven't been integrated into databases
        - Needs to be implemented in systems within health departments and outside health departments as well
    - Model and understand many infections: need to grow workforce
    - More representative data across racial groups
    - Data governance
    - The impact assessments to explore if particular approaches are effective
    - Also missing from electronic laboratory reporting: reason for testing

BREAKOUT SESSION 1

6. What is needed to ensure that the data available are useful for improving forecasting and modeling?
    - Some of these questions were covered in the first session
    - Better to have a discussion as a small group (note: did not use Jamboard during this breakout session)
    - Legal and governance structures and improvements. Models for these structures at various levels.
    - Metadata about data generation process and limitations
    - The ability to disaggregate and reaggregate data based on different variables/characteristics

- Better interactions between those who generate/collect data, modelers, and those who are asking the questions/using model outputs
- Along with metadata, standards on data. A documented agreement on the representation, format, and definition for common data.
- This is a reflection from the opening session—it seems to be legal and regulatory/political standings that need to be transversed and amended. To ensure availability and usefulness, there needs to be front-end work to address barriers/facilitators
- Relates back to governance structure of data. It is problematic to share individual level data due to identifiability concerns. However, presenting aggregated data along singular levels is problematic if we can't break them down along different dynamics. Aggregate data along dimensions should still meet criteria along lack of identifiability.
- What constitutes as identifiable? Should there be ways to do this that doesn't compromise privacy, etc. Is this information that would compromise an individual?
- Statistical re-identification is possible so health departments defensively aggregate data to protect confidentiality. Aggregating is a simple solution to protect confidentiality even though disaggregation is more useful scientifically. Building capacity in people publishing data to know when and how aggregation/suppression is truly necessary to protect confidentiality.

7. What challenges have you experienced accessing USG and other data for your current research?
    - Clarifying comment: USG meaning US government
    - Data isn't available by gender and ethnic and racial background
        - Wanted to stratify analysis and explore gender and racial/ethnic differences
        - Was surprised basic demographic information was unavailable
    - There's difficulty sharing data between the local/jurisdiction level and the USG
        - USG had to scrape aggregate data off local websites
        - Weren't getting local level data in real time
        - Need to improve systems to get data to national level and get granularity that is needed
    - Lack of timeliness in data
    - Current/recent international travel data, geographically disaggregated by place of residence, so we can gauge risk of importation
    - Failure to provide easy access to historical data sets, e.g. NREVSS (National Respiratory and Enteric Virus Surveillance System) data
    - Disaggregated population denominators—granular by age in years, detailed race/ethnic group, primary language, census tract of residence
    - During COVID, lack of standardization of how/what data is collected by different states, how it is made available, where it is made available, etc.
    - USG data can come from more than Public Health—input on data from other sectors is also a consideration
    - Detailed ethnicity data. Usually you have ~5 race/ethnic categories (white, black, Asian, Latino, other) and lots of missingness, and in a diverse place like NYC often that's not granular enough for communicating to the public/modeling what's happening in different communities
    - We did an analysis that required denominator data by single year-of-age, but most denominator data are aggregated into 5-year age bins, so this goes along with the need to disaggregate both numerator and denominator data to suit the analytic purpose
    - You need to be facile with the census website to do aggregates, some of the data can be suppressed
        - Facilitator: The more granular the better?
        - Participant: Yes.

8. How should modeling be used to inform the design and interpretation of epidemiologic studies and surveillance systems?
    - With SARS COV2 (for genomic data) there has been a breakdown between connecting local data and gaps were created between genomic and epi data
    - Can we clarify the question?
        - Facilitator: As Dr. Jernigan was discussing, the proposed Center will model and forecast public health issues and share this information in real time—as part of response, we need to consider epidemiological studies and surveillance systems
    - If there is consensus about specific data that a modeler would need, ex: needing serologic data or ways of designing systems that are all encompassing
        - Modeling systems that are less susceptible to bias—examples: sentinel or sero surveillance systems or systems that use existing resources to conduct surveillance
        - If the modelers need that to make predictions or forecasting, that would suggest or would lead to the allocation of resources to develop those systems
        - In public health the most basic thing is to improve existing systems and routine surveillance for diseases and outbreaks, but if you are going to supplement this with additional systems that will help with forecasting, that's a dialogue that could be opened between scientists doing forecasting and modeling and the folks that allocate resources to the design of these surveillance systems
    - On a real-time basis during a pandemic or outbreak, modeling can be improved prospectively by the epi data and, vice versa, what the modeling is showing on a real-time basis can improve data being collected
        - Identification of gaps in real time and data that fill these gaps—this is a cycle
    - Increase awareness of decision makers and public health experts about what kind of models are available
        - Can apply these models to strengthen the existing surveillance system
    - Experience at beginning of pandemic—the work that Dr. Lipsitch's group did at the beginning of the pandemic in NYC (https://rebeccakahn.shinyapps.io/COVID19/) helped create understanding based on predicted importations of the virus into NYC (participant's group do these detections because they didn't have capability due to testing limitations)
        - The work that Dr. Lipsitch did was very useful in helping develop prediction for when the pandemic might cause community wide transmission in NYC that would be detectable
        - Wanted to implement system for sentinel surveillance and consider remnant specimens (SARS COV2 specimens across the city)
        - Having modeling inputs could trigger other ways of looking for something that might be about to happen
    - Modeling is essential for identifying key data to collect—it can help prioritize data to collect that may not always be thought of as an essential outcome
    - Examples in own work include designing models/studies to understand typhoid and vaccine impact. It was key to understand the role of chronic carriers in transmission, which may not necessarily be something you think of when understanding impact of vaccination.
    - Analyzing serology (antibody) testing/data—"reason for testing" was necessary for interpreting serology data from passive surveillance but not routinely collected in electronic laboratory reporting
    - Can inform sample size estimation to determine anticipated impact sizes
    - Interpreting impact and effect of interventions
    - Additional variables regarding social determinants that can assist with tailoring public messaging (i.e. factors that impact vaccine hesitancy)

9. What structure(s) are necessary to support forecasting and modeling to ensure value is added to the existing research agenda and community?
    - Good question; one of the structures that is needed is some way to support ongoing dialogue between people who are doing surveillance and epi on the frontlines, at the local, state, and national levels, and with forecasters and modelers
        - Dialogue is key
    - Incorporating into curricula for public health professionals and for other professionals that will interact with public health modeling, we need to ensure that people who might be in position of funding or incorporating outcomes have basic understanding of what modeling can do/how it can be useful and also limitations
        - There are many smart people who just see limitations of modeling and the assumptions that are made and dismiss it on the face
        - If we want to ensure these models are pulled into the research agenda and in communities, we need to ensure these individuals have basic understanding of modeling and forecasting
    - Having ongoing evaluation and feedback on which assumptions are correct and which assumptions are incorrect—sharing this with modeling and forecasters would be very useful
        - Create ongoing knowledge of those consuming outputs, such as what's reliable, where are the big uncertainties, etc.
        - Difficult to use output if you don't know
        - These kinds of evaluations need to be transparent and communicated
        - Assumptions in forecast need to be communicated to stakeholder—there was abundance of modelers doing work during COVID, but they came out with different predictions and people dismissed it or didn't know how to use it
        - There's a big need to evaluate the outcome against what actually happened, understand how this might have happened, and how this can be used to interpret current outputs so that the whole enterprise isn't dismissed but rather evaluated and understood so that decision-making can be appropriate in its response
    - Clarification—notion of continual improvement: evaluate what happened with predictions and use evaluation to improve on future analytics
        - Participant: yes, exactly—and communicate this to those using the outputs
    - Agreement expressed for the above comment, important to present evaluations
        - However, people are more likely to present positive evaluations
        - People are less likely to present the evaluations if assumptions are incorrect
        - It's important to present both positives and limitations/challenges and take them into consideration for future forecasts
    - What's an assumption that's made in a forecast that should be communicated to end users?
        - Could be general, i.e. in this model we've assumed the proportion of x is the same across the entire population OR the likelihood of behaving in a certain way is the same OR it might occur within this probability range
            - Recognize this isn't a specific example
            - These examples explain that perhaps in a model it doesn't necessarily account for all the diversity (stochasticity) that might exist in a real population
            - General assumption that needs to be communicated
    - Additional training programs to develop background/literacy around modeling, limitations, inputs among non-modelers
    - Politicians making demands/unreasonable demands regarding questions and responses to answers (e.g. politicians may not listen to answers demonstrating uncertainty)—how to handle? Would like training on this.
        - Who are the go-betweens between modelers and politicians?

- Training *for* politicians—if the onus of training is on the modelers or on health department staff, that feels untenable
- Was thinking more trusted informants in the political sphere

- Interested to hear if anyone is thinking about structure that would help build/go beyond health-related data/different sectors?
  - Data systems management—how to collect data and make it available with appropriate caveats to researchers? Data during the pandemic was collected from non-government sources (Facebook, non-traditional providers, etc.). If limitations are not well understood or easily available to those who'd like to use it, there's a role there to help facilitate that.
- Something that may have gotten reduced in the most recent census is primary language spoken in the household—this was a key variable that helped us identify vulnerability in our local community before but removed the granularity there. In defining communities at risk, this was a step backward.

10. What are key interfaces between modeling/forecasting and genomic surveillance of pathogens?
    - Huge problem currently—genomic surveillance is obviously of interest and new capacity is being generated
      - Genomic surveillance is currently conducted by academic labs, public health labs, by some commercial labs, etc.
      - Often labs conducting this surveillance experience issues related to selection of samples for genomics
        - Ideally you would have an unbiased selection
      - First issue: There's no metadata around selection, which is problematic
        - In some situations, you have it, in others you don't such as with GISAID (repository for genomic sequences for SARS COV2)—there is no information on why particular sample was sequenced
      - Second issue: connecting genomic surveillance data with demographic or epi data—huge challenge currently
        - Also difficult to get genomic data into other places such as health departments, surveillance systems, and other places where data is aggregated
      - So many issues related to this—to provide meaningful input and to model, we need to understand source, selection criteria, and be able to link it to other data
      - These areas require a lot of investment in the US
    - Time/temporal component—took considerable amount of time during pandemic to start or even complete sequencing that could be helpful and this could/can help us learn about how virus is changing in real time
      - Be ready from the start to work on sequencing on a representative scale and maintaining this throughout
      - Support expressed for comment on data and representativeness
    - Have there been instances where this has worked well (regarding temporal aspect)?
      - Thinking about this from high level across pandemic
      - In Minnesota, on smaller scale, we have used genomic data to put together clusters of how individual transmission events have been related to each other—this process is used commonly
        - Minnesota is a place where it is working well
      - It is working well, ramped up quickly, and worked with partners to get theoretically randomly sampled specimens to sequence in addition to doing targeting sequences for vaccine breakthroughs and seemingly dynamic outbreaks like in nursing homes

- Working well in NYC, labs doing random sampling, and providing data to public health
  - There is a time lag there and it needs to be expanded further
- Need better methods in general that should be disseminated to detect early emergence of viruses that have particular characteristics that are concerning
  - There are characteristics of viruses and mutations that are concerning (consider transitions in genotypes that affect transmissibility, etc.)

11. What are specific recommendations for the integration of verification and validation processes to ensure quality results?

- What kinds of social information can be provided? And by what institutions or organizations?
  - What kinds of data collection tools have been used? Have they used standard or adopted questions?
  - What kind of information is being collected—quantitative, qualitative, mixed method?
  - What is the sample size?
  - Communicating information about how end result was achieved should be highlighted
  - Particular community, gender, zip code, etc.—that all should be reported

General comments at end of first breakout discussion:

- These are really important topics—in NYC two modeling groups were involved (representing the city and at the state/national levels) and as a person who was involved in the surveillance and generation of data, the conversations were extremely useful
  - Had twice-a-week conversations with modelers
  - Was difficult to interpret outputs when they were differing
  - Center is a good idea and is needed
  - Participant feels excited about what will come out of work and how it can improve response activities

12. What structure or design would best achieve the desired scope including the coordination and collaboration with the STLT, academic, health care, private industry, and interagency organizations?

- As much as we want to acquire new data, there's a tremendous amount of data that's already there but we have problems accessing it. So the low hanging fruit is providing access to data that is already there (like EMR). The EMR data is already sitting in the system, it's tremendously useful, has low latency, is highly specific, and the obstacles to get it are mostly jurisdictional and privacy challenges. The solution could be a federated surveillance—we don't need to transfer rich medical records from holders (hospitals, etc.) to a central place—we just need to allow/enable querying of the data so that we can get the counts/numbers.
  - The system we would like to see in the coming years is an API (rather than a national system) that all providers/holders of EMR data will agree to and that they can be queried as needed to track different phenomena
  - There are 100,000 outpatient clinics in the country, but there are only 100+ health information exchanges that are doing a heroic job of collecting data, interfacing to their local outpatient/inpatient systems
  - We should put in place a national system that interfaces to these health information exchanges and queries them so that identifiable data should never leave state boundaries, but with API interface we could get real-time access to it
- Follow-up question: when you talk about the API pieces, are you thinking about bulk FHIR?
  - RR: exactly thinking of FHIR and agreeing with HIEs for an interface that would allow the national center to query all of them with the same query. The query types can be preapproved through some type of agreement so that you can't query everything (some privacy thresholds in place in regard to small counts or identifier. You can also use a

technology called homomorphic aggregation (saying we can add up accounts from everyone without actually knowing who one of the contributors is). We can count without disclosing information for a specific hospital.
    - The data is already there, the only thing that is missing is the interface and the agreement to use it
- Biggest concern is the audit or provenance of the data (and metadata about the information we get) is incredibly important. Usually, we're dealing with aggregate accounts that are coming from multiple sources (outpatient clinic testing, hospitalizations, contact tracing study). Particularly at critical times in an epidemic, that provenance can have a huge impact on what that data actually means.
    - Having more standards/standardization on how that provenance is tracked and making sure it flows through to the end consumer of the data could make a big difference in ability to effectively use that information and deal with changes that happen over time because there are very quick reporting changes
    - It would give the ability to have a more standard way to deal with the abrupt changes
- Could we add long-term access to mobility and contact data (Google, Safegraph, Cuebiq, etc.) to the list?
    - That has become very valuable during the pandemic, but will probably go away after the pandemic unless we do something about it
- We haven't really talked about organizational coordination -very big problem. There is a lot of talk about a National Center for modeling and forecasting infection diseases—what will this do for local jurisdictions? What does that mean for private industry? What are the interagency links that will allow more effective delivery of information that is important at local level. This will take time and investment/personnel investment.
    - One particular way that we should handle this is to learn what has been done with weather forecasting; element of international collaboration for sharing data in real time; and integration. Took many decades to set up. There is an element of international incorporation collecting the raw data. Great example of collaboration across the world.
    - There's also an element of national modeling and integration and authentication; creating authentication and putting approval in the nation weather service. This is not at the professional level.
    - There's another element at local value added, local modeling, and local communication—local people translate data for local consumption

13. What are the lessons or successes from other models/approaches, e.g. Models of Infectious Disease Agent Study (MIDAS) or Research and Policy for Infectious Disease Dynamics (RAPIDD)?
    - RAPIDD is no more and MIDAS is in a different form now in that it's more voluntary and inclusive. They were considerably advanced. The communities of infectious disease modeling. Taking it from a place where it was very disaggregated to a place where it's much more data-driven.
        - Led to a lot of the issues we talk about now; credit to those people for bringing it together; people were coming together, and they had resources to support their work and it was something of recognized value
        - We're not in the same place, but there's a need for infectious disease modeling, forecasting, projection. As people will complain about weather predictions, they're going to complain about inaccuracies of these models; but there's an expectation at government and institutional levels to use these tools/models so here to stay.
        - How do we identify the central issues that are most important for advancing the field? Sometimes we get sidetracked into things that may not be as meaningful. Find the directions that we really need to move in.

- Participant was involved in MIDAS; view it as a scientific exercise (aimed at understanding what is happening), but we need for a technological initiative (creating specific capabilities). We need to create technology exercises. Example: flu forecasting challenge by CDC is great example; weather forecasting includes tech aspect as well.
  - New Center must focus on technology since we're way behind
  - Agreement with distinction between technology and science; also need operational structure to make it work in a sensible way
- Happy to say a few words about what we think worked well in the structure of RAPIDD
  - We need to have group projects to advance the field; what helps with modeling forecasting. Maybe we should have working groups?
  - There is a need for research into models and what worked well in the field
  - Royal Society fellowship in UK had freedom to do what they thought was useful in the field
  - Flexibility around meetings or doing workshops is key
    - Another participant expressed complete agreement—like the distinction between science and technology—we need investment in operational ID analytics, modeling, and forecasting

14. Who are the stakeholders that are critical in the implementation of solutions?
    - Media
    - Emergency room clinicians
    - Local public health officials with meaningful audience/linkage with governing officials/ politicians in power
    - Owners/custodians of EMR
    - Local public health officials and practitioners
    - Hospital administrators
    - Funding agencies!
    - Federal health authorities
    - The lay public? At least on the behavioral side.
    - Data warehouse companies and other data generators
    - Scientists (modelers, immunologists, virologists, behavioral scientists, epidemiologists)

15. What are other important gaps that need to be considered and addressed?
    - Data access (existing and novel streams)
    - Integrated platforms for data curating, sharing, analytics, modeling, forecasting
    - Development of operational systems/centers for modeling/forecasting with archiving, quality control, etc.
    - Incentives for EMR custodians
    - Unified API for querying rich data to produce counts
    - Understanding in the broad range of stakeholders of capabilities and limitations of modeling/ forecasting/etc.
    - Maintenance of capacity with a practice focus
    - How do we measure success? A job well done in modeling?
    - New technologies to sense and capture disease markers
    - Global capacity building in modeling

BREAKOUT SESSION 2

16. What is needed to support and enhance the understanding and use of modeling, forecasting and sophisticated analytics among the federal, STLT, and global public health workforce?
    - Communication is the absolute key, first and foremost what has to be done, stakeholders and users need to talk to modelers and forecasters and need to learn from one another; must understand perspectives so that they can each inform one another going forward
        - Takes time and investment
        - We've already seen the dividends, in terms of the integration of modeling and analytics in the 2009 flu pandemic—appetite for it has been enormous
            - Changed because of continued investment and communication
        - Both passive and active attempts make it better
        - Critical to understand, interpret, and use forecasting and modeling correctly (modelers, forecasters, and end users)—need to understand strengths and weaknesses, what it can and can't be used for
        - Use it in an intelligent fashion despite its limitations
    - Above comment speaks to need for broad-based training in modeling (not necessarily because we need more modelers but rather to broaden the workforce generally) and need for more informed consumers of modeling to understand how models are generated, what are underlying data needs, and what are essential limitations and uses of models
    - Participant expressed agreement for comments so far; key is sustained relationships that aren't only formed ad hoc in a crisis but are continuously developed
        - Participant received many panicked emails from groups and organizations at start of pandemic
        - When relationships are formed, there is a big learning process in the development of trust
        - All these things are happening in real time in a crisis—having ways to build relationships, building trust, and conducting conversations outside of a crisis will make us savvier consumers
        - Will also help with understanding everyone's capacity
        - Resounding support from group
    - Clarifying question from observer: is there guidance, or a set of materials, tools and resources, or positive examples of what this looks like (i.e. if a public health department responds rapidly but wants to have appropriate caveats in place)?
        - Relationship building and mutual understanding and support makes sense, but what does that look like?
        - Participant can offer two examples, one in crisis and one in peacetime
            - Crisis example: pandemic started in January 2020—NY Dept of Health and Hygiene reached out to participant's group to sign an MOU and DOA for forecasting and modeling of new pathogen in NYC
            - Got logistics out of the way early on, then 6 months later, funding was provided
            - This is reactive—happened to be a situation where there was a large entity where they got some logistics out of the way early on
            - Peacetime example: CDC flu forecasting effort has allowed communication to flourish, what are the problems and issues at play, what is the science of technology and data access—all these questions have been addressed in the course of this effort
            - Participant was heavily involved in a forecasting leadership role in this process—saw firsthand the reality of communication over time and the benefit of sharing operational forecasts, continuing the work in a sustainable way, etc.
                - CDC saw the importance of forecasting and then set up a forecasting competition
                - There wasn't funding initially, which was a sore point

        - This process brought CDC into contact with forecasters and modelers in a more intimate way
        - Two and a half years ago, CDC also hosted a mock forecasting exercise and simulated a pandemic (had mock black ops team, press conferences, fake press corps, simulated vaccine shortages, etc.)
            - This participant participated in and observed this process
            - Degree to which modeling and forecasting was involved represented a radical departure from traditional practice because of the function of the time spent before (in the previously mentioned peacetime project) demonstrating the benefits of these processes
            - Allowed CDC to see that there was merit and opportunity in these tools, and they needed to be brought into decision infrastructure
    - Another participant had a similar experience to the previous participant in the state of California, which was moderately successful and eventual funding was helpful
        - Where the flu forecasting could have done better: had we somehow directly clicked over into the forecast hub and other efforts, we would have brought different teams together (which is important)
        - Part of building trust and understanding is getting people to see different models, approaches, and thinking processes in the same place
        - Confusion and mistrust around the number of available models
        - Created a situation where people in various parts of government, etc., were overwhelmed with options
        - We need room for innovation and some structure
        - We also need some mechanisms for centralization and build out established communication and trust
        - Need to look at what predictions and forecasts look like next to one another
    - Need for impartial outsiders who are not involved in modeling themselves who can help to interpret and, when presented with multiple models, can evaluate them for limitations and use
        - Have a person to distill formally and informally aggregate all information coming from different forecasts

17. What is working well to support the use of modeling and forecasting for public health? Are there things that aren't working that should be addressed or avoided?

    a. Are there things that aren't working that should be addressed or avoided?

    - Support expressed for comment in first session; ahead of time, identify individuals at a public health agency who have background in modeling, even if they aren't actively working on it, who can serve in these roles
        - There aren't many people who understand modeling methodology and who should be involved
        - Identifying people in agencies or groups who can provide this context would be helpful
        - Tap into resources that currently exist until we can increase knowledge and understanding about modeling and forecasting
    - Points back to training need to have individuals who are literate in modeling
    - Not everyone who's doing modeling has right expertise or input for the moment
        - In a crisis situation, you want to identify the people who have the ability to provide insight and aren't just providing dogma
        - Difficulty in this work
        - Want operational center stood up in peacetime (i.e. National Weather Service) that provides model and predictions in an operational way

- Not to say that this Center won't be without issue—but we can continue to evolve the structure where there's a centralized and operational place that is doing the day-to-day work of forecasting, that has the ensembles built out, that is using multimodal approaches, that is having everything regularized to understanding that this is a scenario, this is a true forecast, etc.
- Need predictive capacity
- Won't be a linear process—there will be hiccups along the way and there are examples of success led by other participants on this call
- These are steps along the way in how we learn how to do something—the more we do in peacetime the more success we will have in the long run

- Facilitator: Is there anything that is working well?
  - CDC has done some work on it and has begun to fund it
    - MIDAS and RAPIDD were really important for training a whole new generation of people to think about this
    - Established cadre of individuals who use suit of mathematical tools to understand infectious disease problems
    - Much of work of forecasting isn't scientific, it's technical
    - There are scientific questions that need to be addressed, but it needs to be moved into an operational setting
    - What CDC has done with seasonal flu forecasting has been a transitional step toward this
    - Relies on many people working for many years without funding support during in-between times that doesn't work well at the interface of government and academia
    - Need sustainable funding for viability
- Observer: The above comment stimulates thoughts about the Center and workforce training—Center would help socialize the field around what data are, how you talk about it, how you contextualize it—are there other means to this same end, in addition to training, that the Center could do to support what JS was describing?
  - If we are talking about the Center as a proposed Center for modeling and analytics, if it's going to be modeled after what's been done, look to existing models
    - National Weather Service
    - National Service for Environmental Prediction
      - Both provide operational use of models
      - Run models in day-to-day with experts who run post-processing and interpretation because that improves the accuracy of the models in addition to communicating and distributing it
    - UCAR and NCAR
      - Doing research on operational forecasting
      - Where do we better use satellite data? How do we glean data?
    - Academic institutions
      - train people and place them in post-docs at UCARR and NCARR
    - There's an ecosystem that exists that marries the science and operational aspects of it
    - Operations are key
  - Support for the comment above
    - Agree, you need all these components in terms of people who are employed to do the day-to-day operational work as well as the go-betweens (between the academics doing research and the operational people) as well as the support for the academic research environment to push the state of the art
    - Weather forecasting provides a model for this
    - Must also understand how we deal with and communicate uncertainty

- Support expressed for the above comment, would have been at top of his list 10 years ago
  - Has been pleasantly surprised over the course of the pandemic to see that the reality is that this doesn't always play out this way (i.e. there isn't a military person demanding numbers from a researcher with certainty during a crisis)
  - Feels military is habituated to uncertainty, same with businesses
  - Trying to scope out possibilities of what is coming down the pipeline
  - People are becoming comfortable with uncertainty
- This is a bit of a tangent, but you can't do anything to change the weather forecast, but in infectious disease forecasting, you can intervene (through specific public health actions) and this can change the outcome that's predicted
  - Behavioral component should also be considered as factor that can change outcomes
  - This is an area where there's a need for improvement in how we communicate limitation and how we understand and learn to incorporate feedback in the model itself
- This question has been posed all the time—what if we make a forecast, people respond, and it's wrong?
  - That's good—would use a model to show why this happened and that's okay
  - Point of contention: system is nonstationary in statistical sense during pandemic
  - There are many unknowns about the future that can influence the trajectory and case counts and hospitalizations in the future
  - When JL put together his hub, he called it a "scenario modeling hub"
    - Was clear to state that we are making projections and not forecasts; distinction is important
    - Forecasts are meant for calibration and accuracy in stationary systems when people react in aggregate (example: seasonal flu)
    - Behavior during this pandemic is important and must be considered and recognized

18. What is needed to support training for decision makers and frontline staff to improve public health decision-making?
    - Authority
      - Public health officials over the course of the pandemic have been politically disenfranchised
      - Begs the question: is there a reason to do the science?
    - Effective public communication
    - Better understanding of the types of questions for which modeling may be useful (or not useful)
      - Decision makers and frontline staff
      - Or understanding when a model is useful
    - Complete lab and epi data
      - Often get data coming in through diagnostic labs or through case report forms that are missing patient contact info, race/ethnicity, etc., which are critical to understanding data and making decisions
    - Established networks of communication and collaboration across those who collect data, those who interpreted it and generate models and the decision makers themselves
    - Epidemiologists (or individuals with knowledge of epidemiology and data) in positions of communication with policymakers. Need to advocate for public reporting of meaningful data.
    - Trust in the decisions by the public, which speaks to the need for better communication.

19. What are the lessons or successes from other models/approaches, e.g. Models of Infectious Disease Agent Study (MIDAS) or Research and Policy for Infectious Disease Dynamics (RAPIDD)?
    - Observer: have you seen models of training or substance/content of training that would be seen as high priority
        - An earlier participant mentioned MIDAS and RAPIDD—would these models be helpful?
            - Yes, these models just helped stand up the field
    - Workshops in which modelers and frontline public health officials come together to learn from one another
        - The partnerships don't have to be workshops, per say
        - Brought in people from public health agencies, CDC, CSTE, White House, etc., as part of first forecasting grant (2012)
            - Goal was to understand what each person saw, what they could do with it
            - Got the conversation rolling
        - DOD has also proved to be a good model
        - Flu forecasting model has been a great example of a workshop that has gotten everyone involved
    - Trusted members of staff who have had formal training (e.g. workshops/short courses) in model building and interpretation
    - Education on uncertainty, the current capabilities and limitations of simulation, inference, and forecasting

20. What is needed to improve the communication of forecasting and modeling results and information to decision makers and the public?
    - One of the things needed is continuous and transparent evaluation and feedback about what is working and what isn't (assumptions, how successful were the models and why/why not). decision makers don't know what to make when they get different outputs from different models. Needs to be a way to communicate honestly and transparently that eval into the decision-making arena, as well as plainly as can be so folks can understand.
    - Also distinction between different kinds of modeling and projections (even with seasonal flu). In scenario modeling we think about what would happen if we modeled/intervened this way, what would our trajectory be like? Decision makers may not always understand the difference between different kinds of modeling.
    - We had a situation where two external modeling groups provided two different forecasts. Politicians responded: come back to us when you agree/align on a forecast. Only one unambiguous answer was acceptable to politicians at the time.
    - Communicating limitations and assumptions is important and should be upfront
    - Weather forecasting has come up (international and cooperation framework example)
    - Visualization, probabilistic outputs. There's a study coming out, and could be done more broadly.
    - Agree on communicating outputs visually—what works on communicating complicated data visually to decision makers
    - Some success in using dashboards, decision makers came to decisions quickly. The ability to change assumptions and allow user to have some agency could be useful.
        - e.g. "R Shiny" apps for effective visualizations for scenario exploration that allow users to change assumptions, e.g.: https://rebeccakahn.shinyapps.io/COVID19/
    - Facilitator: do these two groups need different support?
        - Yes, different outputs would need different models. The questions from audiences will be different. The public would want to know "what's happening here where I am?"—how to make them relevant.

21. What is needed to increase two-way communication between modelers and decision makers to better define the highest priority questions for modelers to address?
    - Wherever the data source is coming from, folks could help provide context to the data. Shouldn't be just a dialogue between modelers and decision makers, should include data scientists who understand the data.
    - Not imagining a platform but in NYC we had twice a week conversations with modelers and 1) decision makers and 2) technical staff. These dialogues are important.
    - Calls organized by CDC (four times a week) around forecasts, interventions, hospitals. They were for modelers, but people present there understood decision-making conundrums and that helped bridge the connections between modelers and decision makers
    - Connections with local public health departments were helpful, but not necessarily natural to think about otherwise. How to sustain those relationships in peacetime (outside of the pandemic)?
    - Connections based on personal networks—was asked to help other public health departments who were also trying to form emergency unfunded collaborations. I am concerned about whether this Center would focus on a national level. Will this primarily inform national-level concerns or how will it also address concerns/needs for public health/health departments to form local connections?
    - We were able to get everyone in the room, consistent weekly meetings to provide feedback and input—more of a working session. Pandemic was impetus to start and continue this mindset—being proactive versus reactive. Helped mobilize us to be prepared next time.

22. How should these data (forecasting and modeling results) be shared and are new tools/platforms needed to facilitate?
    - Asked for clarification: what are "these" data?
        - Answered "forecasting and modeling data"
    - Easy access to model assumptions and links to input data
    - Web interface/Shiny App.
        - "R Shiny" apps for effective visualizations for scenario exploration that allow users to change assumptions, e.g.: https://rebeccakahn.shinyapps.io/COVID19/
    - Example: https://covid19scenariomodelinghub.org/viz.html
        - Good example because it has both interactive visualizations and also detailed GitHub
    - Interactive modeling with active results displaying (like a weather radar app)
    - Not really [participant's expertise]—platforms, etc., but do think whatever they are—would be great to connect to public health APIs and or github sites for near real-time updating
    - Define the lowest common denominator tool that small county health groups would have access to that does not require specialized licenses/knowledge and large scale file transfer capabilities such as Google drive
    - The COVID-19 scenario modeling hub meets needs for many audiences because it has detailed and high-level data depending on how much you'd like to dig in
    - Scenario modeling for COVID-19, it would be nice to have something inventive
    - One thing that would be great for visualization is seeing radar/models move through time—how can the user interact with the data as it moves through time with the data, similar to weather forecasts?
    - Like weather displays on apps

23. Who are the stakeholders that are critical in the implementation of solutions?
    - Communication specialists and behavioral scientists
    - Politicians to frame key questions; health departments to describe data generation processes and limitations; modelers regarding methods, assumptions, interpretation
    - Employers—businesses (can implement workforce policy) (important to supply chain)
    - Elected officials—make decisions about policies
    - Health department staff
    - Healthcare organizations
    - Data scientists
    - Coroners/morticians
    - Press media
    - Communities that experience disparately large burdens—and more severe outcomes
    - A big stakeholder includes the public, specifically communities experiencing disparate impact and high burden
    - Healthcare organizations, elected officials, policy makers, health department staff, data scientists,
        - Elected officials must make decisions about policy
    - Private companies/employers—have interest in workforce, supply chain, etc.
        - They can also help with interventions such as vaccination
        - Fatality management—coroners, anyone who must handle fatalities

24. What are the bold, meaningful, and feasible implementation and utilization goals to be accomplished in the next 5 years that were not previously mentioned? (refer to responses to large group discussion Q2)
    - Success of different approaches, models, and forecasts; explore where there are certain successful approaches in certain countries (like developing countries) that are not successful in other countries. If the country has a different healthcare system with a healthcare provision system, some forecasts can be successful in context to certain particular healthcare provision systems. Thinking about a more global level rather than the US only.
        - Examples: some data can't be collected because they don't have capacity to conduct tests; had to send tests to another country to get results in some instances
        - Building capacity across countries

    a. Which goals are the greatest priority and why?
        - Electronic medical records are the biggest untapped resources in this country; UK has national healthcare system and can tap into EMR right away and see real-time data
            - Because the US healthcare system is so fragmented, there is no easy access to electronic medical data. The data is there already, it's just technological, regulatory, and incentive barriers to access it.
        - Linking the modeling of two syndromic surveillance; seeing what happens in ER in real time and then linking that to the research questions and what we're hoping to accomplish

25. What structure or design would best achieve the desired scope including the coordination and collaboration with state, territorial, local, and tribal (STLT) public health entities, academic, health care, private industry, and interagency communities?
    - Probe: collaboration across sectors—what can support design, sharing, and capacity?
    - Collaboration between different hospital systems; nationwide collaborative with key members like hospital systems, hospital associations, health departments to build relationships and have ongoing conversation; maybe like a forum or a meeting to see what's going on across the country; make sure to already have relationships in place so that it is easy to call upon

- Agreement expressed with above comment: It is important to explore what relationships are in place; maybe there is already some type of collaboration between academic and private sector or education—use existing relationships to start work
- If we focus on things that are of real value to the local people, it's much more likely that we could find the human power to do the work and provide the data and give feedback. Should start with goals that are meaningful and valuable locally and will likely lead to being able to better assign resources; if we don't, it'll come across as a burden to people in local areas.

26. Who are the stakeholders that are critical in the implementation of solutions?
    - For developing countries, involve churches and religious leaders since they provide lots of support; Ministries of Health—without their support you can do nothing, decision makers have to give support and train them and help them understand what you are doing
    - Hospital associations; healthcare administration; long-term care agencies, skilled nursing facilities—they are usually left behind so have them be in the conversation up front
    - Start with customers/consumers; suppliers of the raw input/data, whoever owns or has access to the data that we need
        - EMR/EHR—looping them in and finding ways to incentivize

27. What are the important implementation or utilization gaps that need to be considered and addressed?
    - Design a data standard and an API for querying locally held rich data sets and returning counts that can be used for tracking and forecasting
    - Is the tool user friendly? How easy is it to transcribe the data?
        - Explanation of sticky note: is the tool user friendly; how easy is it to transcribe the data; consider if the data is replicable; focusing on how easy it is to share with others especially when you have a quick turnaround
    - Data confidentiality and sharing the data with stakeholders
    - The cost of the tool
    - What tools are currently available and how they can be used in practice
    - Infrastructure (e.g. high speed Internet) to use the tool and access data

28. What kind of staff are needed in the implementation of solutions?
    - Experts in computation, data analysis, and information technology
    - SMEs from the areas under consideration, i.e. infection preventionists, microbiologists. State and local health departments.
    - Data collectors and data analysts

29. Who are the additional stakeholders that are critical in the implementation of solutions? SKIP if short on time.

    NOT ANSWERED IN THE SECOND BREAKOUT SESSION DUE TO TIME CONSTRAINTS.

APPENDIX D

Article "Infrastructure as Social Sensor"

Infrastructure as Social Sensor: The Case for Better Collection and Integration of Infrastructure Data for Improving Pandemic Response

Authors: Daniel Erian Armanios (darmanios@cmu.edu) and Nicola Ritsch (nritsch@andrew.cmu.edu), Department of Engineering & Public Policy, Carnegie Mellon University

There is increasing consensus that infrastructure is crucial for connectivity, and that access to infrastructure is asymmetric. Therefore, tracking the natural buildup of physical infrastructure (and gaps in digital infrastructure) may serve as a crucial means for more rapid pandemic response. Across numerous disciplines, the general consensus is that physical infrastructure enhances connectivity and, in so doing, improves a variety of economic and health outcomes. This finding is corroborated across disciplines such as urban studies,[1-6] medicine,[7,8] anthropology,[9-12] civil engineering,[13-15] economics,[16-28] sociology,[29-33] political science,[34] and management.[5-37] Within the realm of health care in particular, road infrastructure helps better ensure greater healthcare delivery and access. Physical infrastructure not only serves as the conduit through which individuals arrive at healthcare centers, but also enables the delivery of essential medical supplies to said facilities.[38] For example, prior studies have found a correlation between road construction and better medical care and vaccination rates.[39] In some cases in fact, distance and travel time were identified as the most important factor that determined the use of health facilities.[40]

Yet, such access to connectivity is often skewed. For instance, more marginalized communities, either by race or income, find their access to infrastructure more restricted,[34,41] if even available at all.[41] In fact, there is increasing recognition that infrastructure plays a critical role around social determinants of health. As Yancy eloquently noted,[42] being able to maintain social distance, amongst other social activities, are "issues of privilege (emphasis in original)." In fact, the empirical literature plays this out, while marginalized communities are more likely to mask and try to socially distance,[43,44] they still find themselves more likely to frequent "superspreader" locations such as cafes.[45] Space, as enabled and constrained by the built environment, seems a key factor on exposure to pandemic, and a response strategy that leverages this infrastructure data would better target response to those more likely at risk.

How could natural infrastructure buildup (or lack thereof) inform our pandemic response? The enhanced connectivity associated with social distancing in light of the COVID-19 pandemic makes physical infrastructure crucial in health and economic outcomes following the conclusion of the pandemic. Yet, it may become an acute weakness with a pandemic that spreads more in locales with greater social proximity. In fact, some have even argued that physical infrastructure was important in the rapid spread of the coronavirus across and beyond China.[46] At the same time, a lack of infrastructure, especially digital infrastructure such as broadband, may require individuals to leave their homes to frequent cafes and other "superspreader" sites to engage in virtual work, only increasing pandemic exposure for such individuals.[45] Therefore, natural gaps in broadband coverage and natural buildup in road & bridge networks could serve as a crucial way to target contact tracing and economic relief that can mitigate pandemic spread and help improve compliance to stay-at-home public health mandates. Currently, infrastructure is used as a "social autopsy" of those

disconnected and isolated from society, as was done in analyzing who was more likely to perish in Chicago's 1995 heatwave.[47] Perhaps we can use infrastructure in more real time as a "social sensor" that leverages natural buildup and gaps to more directly identify key vulnerabilities. Rather than as a retrospective "autopsy," infrastructure would be transformed to support our abilities to take a "social biopsy" to more quickly, and in real time, identify vulnerable individuals with greater exposure likelihood, thereby increasing pandemic response speed.

More direct collection and integration of built environment data (both physically and digital) from the US DOT, FCC, and others, tied to demographic data from Census, would greatly advance this possibility.

References from article *Infrastructure as Social Sensor: The Case for Better Collection and Integration of Infrastructure Dtata for Improving Pandemic Response*

1 Jacobs, J. *The death and life of great American cities.* (Random House, 1961).

2 Jacobs, J. *The economy of cities.* (Random House, 1969).

3 Tzoulas, K. *et al.* Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review. *Landscape and Urban Planning* **81**, 167-178 (2007).

4 Hitchner, R. B. in *Highways, byways, and road systems in the pre-modern world* (eds Susan E. Alcock, John P. Bodel, & Richard J. A. Talbert) 222-234 (Wiley-Blackwell, 2012).

5 Fischer, J. M. & Amekudzi, A. Quality of Life, Sustainable Civil Infrastructure, and Sustainable Development: Strategically Expanding Choice. *J Urban Plan Dev* **137**, 39-48 (2011).

6 Chen, C. L. & Vickerman, R. Can transport infrastructure change regions' economic fortunes? Some evidence from Europe and China. *Reg Stud* **51**, 144-160 (2017).

7 Pathak, E. B., Reader, S., Tanner, J. P. & Casper, M. L. Spatial clustering of non-transported cardiac decedents: the results of a point pattern analysis and an inquiry into social environmental correlates. *Int J Health Geogr* **10** (2011).

8 Luxon, L. Infrastructure – the key to healthcare improvement. *Future Healthcare Journal* **2**, 4-7 (2015).

9 Moser, C. O. N. The asset vulnerability framework: Reassessing urban poverty reduction strategies. *World Dev* **26**, 1-19 (1998).

10 Anand, N., Gupta, A. & Appel, H. *The promise of infrastructure.* (Duke University Press, 2018).

11 Appel, H. C. Walls and white elephants: Oil extraction, responsibility, and infrastructural violence in Equatorial Guinea. *Ethnography* **13**, 439-465 (2012).

12 Howe, C. *et al.* Paradoxical Infrastructures: Ruins, Retrofit, and Risk. *Sci Technol Hum Val* **41**, 547-565 (2016).

13 Carlsson, R., Otto, A. & Hall, J. W. The role of infrastructure in macroeconomic growth theories. *Civ Eng Environ Syst* **30**, 263-273 (2013).

14 Faoziyah, U. Who Benefits? The case of the Suramadu Bridge construction. *Procedia—Social and Behavioral Sciences* **227**, 60-69, doi:10.1016/j.sbspro.2016.06.043 (2016).

15 Desai, J. D. & Armanios, D. E. What Can't be Cured Must be Endured: Understanding Bridge Systems as Institutional Relics. *Journal of Infrastructure Systems* **24**, 04018032 (2018).

16 Audretsch, D. B., Heger, D. & Veith, T. Infrastructure and entrepreneurship. *Small Bus Econ* **44**, 219-230 (2015).

17 Acemoglu, D., Garcia-Jimeno, C. & Robinson, J. A. State Capacity and Economic Development: A Network Approach. *Am Econ Rev* **105**, 2364-2409 (2015).

18 Acemoglu, D., Moscona, J. & Robinson, J. A. State Capacity and American Technology: Evidence from the Nineteenth Century. *Am Econ Rev* **106**, 61-67 (2016).

19 Démurger, S. Infrastructure development and economic growth: An explanation for regional disparities in China? *J Comp Econ* **29**, 95-117 (2001).

20 Munnell, A. H. Policy Watch—Infrastructure Investment and Economic-Growth. *J Econ Perspect* **6**, 189-198 (1992).

21 Munnell, A. H. & Cook, L. M. How does public infrastructure affect regional economic performance? *New England Economic Review* **Sep**, 11-33 (1990).

22 Czernich, N., Falck, O., Kretschmer, T. & Woessmann, L. Broadband Infrastructure and Economic Growth. *Econ J* **121**, 505-532 (2011).

23 Esfahani, H. S. & Ramirez, M. T. Institutions, infrastructure, and economic growth. *J Dev Econ* **70**, 443-477 (2003).

24 Banerjee, A., Duflo, E. & Qian, N. On the Road: Access to Transportation Infrastructure and Economic Growth in China. *J Dev Econ* **102442** (2020).

25 Roller, L. H. & Waverman, L. Telecommunications infrastructure and economic development: A simultaneous approach. *Am Econ Rev* **91**, 909-923 (2001).

26 Ananat, E. O. The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality. *Am Econ J-Appl Econ* **3**, 34-66 (2011).

27 Ashraf, N., Glaeser, E. L. & Ponzetto, G. A. M. Infrastructure, Incentives, and Institutions. *Am Econ Rev* **106**, 77-82 (2016).

28 Glaeser, E. L. & Ponzetto, G. A. M. Did the death of distance Hurt Detroit and Help New York? *NBER Working Papers Series* (2007).

29 Samila, S. & Sorenson, O. Community and Capital in Entrepreneurship and Economic Growth. *Am Sociol Rev* **82**, 770-795 (2017).

30 Star, S. L. The ethnography of infrastructure. *Am Behav Sci* **43**, 377-391 (1999).

31 Star, S. L. & Bowker, G. C. in *Handbook of new media: social shaping and social consequences of ICTs* (eds Leah A. Lievrouw & Sonia M. Livingstone) 230-245 (SAGE Publications, 2006).

32 Grannis, R. The importance of trivial streets: Residential streets and residential segregation. *Am J Sociol* **103**, 1530-1564 (1998).

33 Dutta, S., Armanios, D. & Desai, J. D. Beyond Spatial Proximity: The Impact of Enhanced Spatial Connectedness from New Bridges on Entrepreneurship. *Organ Sci* **forthcoming** (2021).

34 Winner, L. Do Artifacts Have Politics. *Daedalus* **109**, 121-136 (1980).

35 Van de Ven, A. H. The Development of an Infrastructure for Entrepreneurship. *J Bus Venturing* **8**, 211-230 (1993).

36 Stenholm, P., Acs, Z. J. & Wuebker, R. Exploring country-level institutional arrangements on the rate and type of entrepreneurial activity. *J Bus Venturing* **28**, 176-193 (2013).

37 Marquis, C. & Raynard, M. Institutional Strategies in Emerging Markets. *Acad Manag Ann* **9**, 291- 335 (2015).

38 Babinard, J. & Roberts, P. Maternal and Child Mortality Development Goals: What Can the Transport Sector Do? *Transport Papers* (2006).

39 Aggarwal, S. *The Long Road to Health: Healthcare Utilization Impacts of a Road Pavement Policy in Rural India.* (Indian School of Business, 2018).

40 Buor, D. Analysing the primacy of distance in the utilization of health services in the Ahafo-Ano South district, Ghana. *Int J Health Plan M* **18**, 293-311 (2003).

41 Jones, S. & Armanios, D. Methodological Framework and Feasibility Study to Assess Social Equity Impacts of the Built Environment. *Journal of Construction Engineering and Management* **146**, 05020016 (2020).

42 Yancy, C. W. COVID-19 and African Americans. *JAMA* **323**, 1891-1892 (2020).

43 Hearne, B. N. & Nino, M. D. Understanding How Race, Ethnicity, and Gender Shape Mask- Wearing Adherence During the COVID-19 Pandemic: Evidence from the COVID Impact Survey. *J Racial Ethn Health* (2021).

44 Huang, V. *et al.* Social distancing across vulnerability, race, politics, and employment: How different Americans changed behaviors before and after major COVID-19 policy announcements. *medRxiv,* 2020.2006.2004.20119131, doi:10.1101/2020.06.04.20119131 (2020).

45 Chang, S. *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82-U54, doi:10.1038/s41586-020-2923-3 (2021).

46 Garrett, L. Welcome to the Belt and Road Pandemic. *Foreign Policy* (2020).

47 Klinenberg, E. *Heat wave : a social autopsy of disaster in Chicago.* (University of Chicago Press, 2009).