Annie Fine, MD

Colead, Epidemiology and Data Unit, Surveillance and Epidemiology
ICS COVID-19 Response, NYC Department of Health and Mental Hygiene

Testimony: Health Information Technology Advisory Committee, ONC

Data needs for high consequence public health emergencies

It is my pleasure and honor to speak today with the committee. Thank you for inviting me to do so. I am a medical epidemiologist and a pediatrician. I have worked in the Bureau of Communicable Disease, in the Division of Disease Control at the New York City Department of Health for almost 25 years and have participated in the management of many large public health emergencies, including West Nile, Anthrax, pandemic influenza, Zika, Ebola, and most recently COVID-19. Since the beginning of the pandemic, I have served as a lead for the epidemiology and data unit within the ICS response.

The importance of having rapidly available and accurate information, and information that is comparable across jurisdictions during public health emergencies cannot be overstated. The COVID-19 pandemic laid bare some of the vulnerabilities and weakness of our public health data infrastructure. The information we have been able to generate, despite massive effort and heroic dedication of labor on the part of public health practitioners, has been slower in arriving than we might have wanted, inconsistent across local, state and federal levels and across geographic regions, confusing to the public and has in some cases failed to answer key questions in a timely fashion. My presentation today will focus on the challenges we have faced in responding to the pandemic and past emergencies, on lessons we have learned and on actions we can take to advance our ability to modify and fortify our systems to manage future public health crises which will inevitably arise and present new and unexpected challenges.

Challenges in Data Collection, Processing and Analysis During COVID-19 Pandemic

Some of the challenges posed by the pandemic are similar to past emergencies. In any crisis, there is a sudden need for rapidly available, accurate and comparable data across many levels of government – local, state and federal. The COVID-19 pandemic also had unique features – first, the absolutely unprecedented volume of data that public health data systems were expected to handle. At the New York City Department of Health, prior to the pandemic, we received approximately 5,000 electronic laboratory reports per day. During the pandemic, we have received more than 100,000 lab reports per day. Our data processing systems and surveillance database were simply not set up to process this volume of data and have strained under the load. A substantial number of hours have been spent working on shoring up these systems (adding servers, improving automation of scheduled tasks, increasing server space and memory, rewriting and moving code from SAS to SQL, skipping processes such as person, facility and provider matching that proved to be bottlenecks, improving load balancing, preventing some lab results such as negative tests and serologic tests from even being processed in the usual way, consulting with outside vendors to identify possible efficiencies), and there are still limitations to the data processing and surveillance database systems.   We are required to match

incredibly large data sets with millions of records outside of the normal surveillance systems, in order to produce calculations critical to the response, such as percent of positive tests.

Second, there was a need to rapidly onboard new sources of reports – for example, point of care tests performed in outpatient clinics and provider offices. We relied in the past on laboratory reporting from labs – so having to figure out how to train and onboard outpatient providers and clinics to electronic laboratory reporting was a time-consuming challenge during a very stressful time for these institutions. This was in addition to the plethora of pop-up testing sites that emerged to accommodate the demand for rapid testing.

Third, there was a predictable demand for several critical data points that have never been easy for public health to collect and for which there is no automated way to get the data. Public health has relied for many years on methods such as individual case investigation, which includes patient interview, and manual medical record review, to gather these data points. Examples include: a) patient outcome and severity of illness, including whether a patient was hospitalized or died of their illness (it is often difficult to review medical records and/or interview patients even once so tracking individual cases over time until a final outcome is basically not possible); b) treatment information; c) occupation, including whether the person is a health care worker, and d) race/ethnicity, which is often not reported completely via electronic systems, especially from laboratories.

Fourth, critical questions about the clinical and epidemiologic characteristics of the virus arose, but there was no efficient way to rapidly design and deploy simple studies to find answers. Some examples include questions about asymptomatic infection, transmissibility in different settings, the risk of airborne transmission, significance and typical duration of persistently positive nucleic acid amplification test as it pertained to the need for ongoing isolation, accuracy of point of care testing, and the risk and incidence of reinfection, the impact of individual emerging SARS-CoV-2 variants, and questions about the relative effectiveness of interventions such as contact tracing, community mitigation measures and vaccination.

Finally, there has been and continues to be a need to focus on the urgent issue of equity, and to address the unacceptable disparities in the impact of the pandemic on people of color, working people, people who live in poverty and other historically disenfranchised communities. The way we in public health collect and analyze data must be improved to capture data in a more meaningful and accurate way, including how people identify themselves with regard to race and ethnicity, and to enable us to proactively identify inequities – and to target and measure the effectiveness of our efforts to reduce them. It is unconscionable for public health NOT to have this information, not to share it with the affected communities and not to do everything we can to respond and work with leadership and partners to mitigate the impacts and work to prevent future inequitable outcomes.

Strategies for meeting data needs during public health emergencies

It is important that we identify and discuss these challenges now so that we can respond effectively to the next public health emergency. First, in all phases of preparing for and responding to public health emergencies, partnership between the local, state and federal organizations is critical. Federal and state authorities must engage and listen to people who are on the front lines, and decision and policymaking should ideally be collaborative, requiring well-coordinated and frequent communication. Second, our public health data systems at all levels must be fortified with the latest technology, that are robust,

tested and able to handle large volumes of data. Many public health departments across the country do not have staff with knowledge or expertise in these areas, and rely upon federal partners for technical assistance and advice on building new systems and leveraging new technologies in the most efficient way possible.

The demand for data to be transmitted from local health departments to state and federal partners is intense, and a huge amount of time is spent by local health department staff in manually entering data into complicated electronic forms, or in mapping data from local systems that do not necessarily harmonize well to federal systems. During emergencies, the primary data points collected and aggregated at the national level should be as simple as possible, clearly defined and standardized across jurisdictions so that they are comparable. This is not a small effort but needs to be undertaken now ahead of another public health emergency. Data points as simple as cases, hospitalizations, and deaths – or more complex ones such as percent positive, require definition and standardization. Reporting should also be done using epidemiologically meaningful data, especially dates. As we have seen during COVID-19, many aggregated data platforms use "report date" to classify cases in time, the date a case was reported by a jurisdiction, whereas using the date of collection of the positive test, or the onset date, will make for more meaningful classification of cases based on when they occur These types of inconsistencies cause confusion among the public and the media, and a loss of credibility for public health.

We also urgently need to develop, disseminate and implement guidance on best practices for the collection, reporting, and analysis of race and ethnicity data, as well as occupation data, so that our front line providers, laboratories and public health practitioners routinely collect this information even outside of an emergency. As we move to more automated collection and processing of electronic data for public health surveillance and response, it is essential that data collected at the time of the patient encounter with the clinical or lab sector be high quality, complete and standardized for transmission into analyzable surveillance databases to provide usable information for detecting and addressing disparities in disease incidence and outcome.

In future emergencies, to quickly understand the threat and the most effective control measures, it might be useful to organize working groups of public health staff from the front lines (local and state health departments and laboratorians), along with federal experts, to rapidly determine what are the most critical epidemiologic questions that need to be answered, and to design quick, implementable but rigorous studies that could be deployed in a few settings to gather answers. Having federal, state and local epidemiologists design relatively simple epidemiologic studies which could be deployed in several locations across the country, with input and ideas coming from all the stakeholders and with rapid data sharing and collaboration on analyses could help to quickly answer some of these questions. An example would be a study such as the convenience testing for SARS-CoV-2 RNA of remnant samples originally submitted for respiratory viral panel testing to determine the prevalence of SARS-CoV-2 circulating in the population at a time when testing capacity was limited and highly restricted. There are so many pitfalls to the interpretation of data collected during emergencies, and an approach like this would serve to generate solid scientifically defensible data. Numerator data by themselves can be extremely misleading and it is critical to use comparison groups where possible for analyses that are relied upon for policy development.

How do we prepare to do this?

First, we must anticipate the types of information that will be needed. We need to be strategic and forward-thinking and build as much automation and interoperability into our systems ahead of time. For any public health emergency, the number of deaths is a critical indicator and therefore, jurisdictions should connect surveillance systems with vital registries for ease of reporting and analysis. Similarly, immunization registries to surveillance systems should be connected. In addition, we need improved person matching and identification. Until there is a national person identifier, it is important to leverage the most current technologies to automate matching. During COVID-19, a huge amount of person time has been spent on manually reconciling duplicate records and on writing code to identify matches. This must change in the future.

Second, we desperately need to improve the linkages between public health data systems and electronic medical records. Important data points such as whether the patient was symptomatic, what level of care was needed, whether hospitalized, severity of illness, length of stay, the impact of therapeutics, presence of chronic underlying conditions, and demographic data such as race/ethnicity and occupation, reside in electronic medical records and should be accessible to public health for epidemiologic purposes. This can be achieved, in part, through advancement of electronic case reporting, but also, the ability to run matches with and/or automated queries of electronic health record aggregate data repositories would vastly improve our ability to understand the spectrum of disease, risk factors for severe outcomes and to better predict and manage the burden on health care facilities. Even something as simple as enabling local public health investigation staff to access patient level data via portals in electronic health records for case investigation, rather than having to request individual medical record retrieval, would enhance efficiency greatly over current methods.

Third, we should work hard ahead of time to develop standardized data definitions and formats for the types of data elements that will predictably be important, including key dates to use when reporting (onset date vs. collection date of a positive specimen vs. date a result is reported to public health vs. date reported from a health department to the public or CDC), race/ethnicity, sexual orientation and gender identity, travel history, occupation, hospitalization). It will take time to achieve consensus, incorporate appropriate data structures and procedures for capturing the data in clinical interfaces used during patient interactions, and to build these standardized packets of questions and data elements into our surveillance systems, but it is doable and will greatly streamline data collection and aggregation in the future. The Council of State and Territorial Epidemiologists and the Centers for Disease Control and Prevention provide an interim case definition as early as possible in the emergency, which is extremely helpful, but we also need to define and provide more granular guidance on the definition of other key concepts that are particular to the condition itself – such as percent positivity, a timeframe after which a case should be counted as a new case, and/or when to count a death as attributable to the condition.

Fourth, we need to add staff to our workforces who are expert at communicating data to the public and the media – to rapidly develop clear visualizations of data that are easy to access so that leadership, the public and the media can understand implications.

If we invest in these endeavors and the personnel needed at the local and state level to implement them, strategically, and ahead of time, and if we can build systems that are automated, interconnected and supported by a workforce that knows how to use them, we will stand far better prepared to cope with the next serious threat to the health of our people and communities.

Thank you so much for your kind attention.