

**Responses by Inder Singh, Kinsa:**

Questions Submitted by Members of the Senate Committee on Commerce, Science, and  
Transportation  
Enlisting Big Data in the Fight Against Coronavirus  
April 9, 2020

## Chairman Wicker

1. Many national and local governments around the world are seeking to use new technology to combat this unprecedented pandemic. Earlier this week, the German government launched an app that allows users to “donate” personal data collected by their fitness trackers or other health devices to help authorities analyze the spread of COVID-19. Authorities in Moscow have launched an app intended to be downloaded by those who test positive for COVID-19. Yet this app raises privacy concerns, as it would allow officials to track residents’ individual movements.

As governments seek to use new technologies in the fight against COVID-19, it is imperative that privacy rights be protected. Are there specific examples of app-based programs you can recommend to policymakers that are both useful in the fight against COVID-19 and respectful of individual privacy rights?

### **Inder Singh Response:**

*Thank you, Chairman Wicker. I believe there is a way to both protect personal privacy and also gather and share the information necessary to detect and effectively respond to outbreaks like COVID-19. The tradeoff between personal information protection and providing information for society’s benefits is a false one. **We can have both.** And we need both if we are going to successfully combat the second and third waves of COVID-19 coming which threaten the fabric of our healthcare system. We must realize that new and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of **immense value to our public health system and our economy.** Data collection, when done right, can **create enormous benefits for individuals and communities, while protecting individual privacy.** We have accomplished that at Kinsa. **Kinsa shares population health insights - the percent of people in a county who are ill.** There is no way to identify an individual from this illness signal. While there are legitimate reasons for sharing de-identified data - for example, for research purposes or to evaluate the effectiveness of a health intervention - restrictions on attempts to re-identify the data should be incorporated into all agreements.<sup>1</sup> This is because with enough effort, even de-identified -- or anonymized -- data has the potential to be re-identified.*

*To reiterate, Kinsa’s population level insights are not personal data, not deidentified data, not even metadata about a person. They are insights about the population. There is absolutely no way to identify an individual based on the percentage of people with fever or symptoms in a county. Our population health insights are available to the public and to public health first-responders at [Healthweather.us](https://www.healthweather.us). We believe this early warning system is critical infrastructure for our country to stop the next outbreak from becoming an epidemic. There is no way to identify*

<sup>1</sup> *We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.*

*an individual from this system. **We need not sacrifice personal privacy for the sake of public benefit. It is possible to successfully do both.***

2. Much of the discussion surrounding the collection of private data to fight the spread of COVID-19 presents two goals – effectiveness and privacy protection – as mutually exclusive factors that need to be balanced. On one side of the balance, it is assumed that greater amounts of personal data, in more granular form, will allow authorities to track the spread of the virus more effectively. On the other side of the balance is protection of individual privacy, which is believed to be threatened by greater surveillance of individuals by the government.

Is this an accurate view of the situation? Are privacy and effectiveness always part of a trade-off, such that the most effective public health measures will come at the expense of privacy, and vice versa? Or do you believe that the most effective policies for combatting COVID-19 can also respect individuals' privacy?

***Inder Singh Response:***

*I believe there is a way to both protect personal privacy and also gather and share the information necessary to detect and effectively respond to outbreaks like COVID-19. We too often focus on the collection of personal data, rather than on building protocols for the protection of personal data; we focus on the private benefits that accrue to companies that monetize data but often do not consider the possibilities of data use associated with public benefit; we focus on the right to privacy, but in doing so, we ignore another important right – the right to information. To be clear, I advocate for personal privacy. I also advocate for saving lives. And it is possible to do both. **At Kinsa, we share real-time population health insights.** Access to timely information is critical when it comes to public health, and the solutions essential for our country's welfare and health need to be advanced. Just as we currently incentivize the creation of drugs, diagnostics and vaccines, **we should be incentivizing the creation, adoption, integration and effective use of novel datasets by local, state and federal public health agencies.** And we can and must also ensure that personal privacy protections also exist. **We can have both.***

## Sen. Thune

3. More and more Americans all throughout the country are turning to online video services to conduct their jobs, education, and social interactions in an effort to practice social distancing. For instance, Zoom Communications had more than 200 million daily users last month. It was found that thousands of Zoom's calls and videos have been exposed to other users online and log-in information has been stolen resulting in many individuals' personal information being compromised.

Did Zoom's privacy policy clearly outline what types of information its platform would collect on individuals? If not, what transparency requirements should be in place for companies like Zoom?

Americans are connecting with each other via online services across all 50 states. Would a patchwork of state laws benefit consumers and better protect their privacy? Should the United States enact a national privacy standard to safeguard consumer's information?

4. Without a federal privacy law in place, the American people must rely on the promises of tech companies that all have varying degrees of commitment to maintain consumers' privacy.

How do we ensure that organizations are actively engaging in data minimization and strategic deletion practices after data is used or transferred?

### **Inder Singh Response:**

*Thank you, Senator Thune. Data collection, when done right, can **create enormous benefits for individuals and communities, while protecting individual privacy**. We too often focus on the collection of personal data, rather than on building protocols for the protection of personal data; we focus on the private benefits that accrue to companies that monetize data but often do not consider the possibilities of data use associated with public benefit; we focus on the right to privacy, but in doing so, we ignore another important right – **the right to information**. To be clear, I advocate for personal privacy. I also advocate for saving lives. **And it is possible to do both**. At Kinsa, we do not share individual data, we share **real-time population health insights**.<sup>2</sup> While there are legitimate reasons for sharing de-identified data - for example, for research purposes or to evaluate the effectiveness of a health intervention - restrictions on attempts to re-identify the data should be incorporated into all agreements.*

<sup>2</sup> *We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.*

*Access to timely information is critical when it comes to public health, and the solutions essential for our country's welfare and health need to be advanced. We must realize that new and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of **immense value to our public health system and our economy**. Just as we currently incentivize the creation of drugs, diagnostics and vaccines, we should be incentivizing the creation, adoption, integration and effective use of novel datasets by local, state and federal public health agencies. And we can and must also ensure that personal privacy protections also exist. **We can have both.***

5. The country of Israel, through its internal security service, has reportedly used smart-phone location based contact tracing to notify citizens via text that they have been in close proximity to someone infected with COVID-19, and ordering them to self-isolate for 14 days. A recent opinion piece in the Scientific American urged democratic governments to quickly follow Israel's lead (see "[As COVID-19 Accelerates, Governments Must Harness Mobile Data to Stop Spread](#)").

Please provide your thoughts on smart-phone location based contact tracing in light of the extraordinary privacy and other civil liberties concerns such an approach raises for U.S. citizens.

***Inder Singh response:***

*We believe you can have both: **preservation of personal privacy and contact tracing to stop the spread of COVID-19**. I recently read a [brilliant comic in Kottke.org](#) that illustrated a contact tracing solution that would both protect personal privacy while also helping stop the spread of COVID-19. To explain briefly: contact tracing apps can work via Bluetooth and regularly send random unique "messages" of letters and numbers to any phones nearby. Because the messages are random and do not use GPS, they contain no personally-identifiable information. Any nearby phones with compatible contact tracing apps pick up and store the random messages sent by others in their vicinity. If any user tests positive for COVID-19, they or their doctor can input that information into the app. Any phones that hold a relevant "message" from the time the infected person would have been near them are notified that they were exposed. Throughout this process, no personally identifiable information has been captured, stored or shared. These are the types of technologies and innovations our country needs, which both protect personal privacy while also giving us the information needed to stop the spread of illness and save lives. The tradeoff between personal information protection and providing information for society's benefits is a false one. **We can have both.***

According to the [Wall Street Journal](#), MIT is developing a contact tracing app for COVID-19 patients and others who have not been infected by COVID 19 that can be voluntarily downloaded to a person's smart-phone. Please provide your views on this approach to contact tracing.

**Inder Singh Response:**

*Assuming that MIT is leveraging the kinds of approaches and technologies that were described in the [comic in Kottke.org](#) that illustrates a contact tracing solution that would both protect personal privacy while also helping stop the spread of COVID-19, I believe this could be a good solution. In order to be effective, it would need to be adopted by a sufficiently large group of people. The solution [Apple and Google are developing together](#) would perhaps have the "muscle" of adoption and might be a better pathway to ensuring effective contact tracing as long as they too are ensuring the protection of personal privacy.*

6. COVID-19 has caused private companies to seek out and utilize health data in an effort to protect users, employees, and the general public from the spread of the virus. Both Apple and Alphabet have released websites to help users self-screen for exposure to COVID-19. This data will be used to help public health officials. However, these tools also allow technology companies access to user's health information which the companies could in turn profit from in the future.

How are technology companies balancing the need for timely and robust reporting to prevent the spread of the virus with the confidentiality and privacy of the participants?

**Inder Singh Response:**

*Data collection, when done right, can create enormous benefits for individuals and communities, while protecting individual privacy. **We can have both.** We too often focus on the collection of personal data, rather than on building protocols for the protection of personal data; we focus on the private benefits that accrue to companies that monetize data but often do not consider the possibilities of data use associated with public benefit; we focus on the right to privacy, but in doing so, we ignore another important right – **the right to information.** To be clear, I advocate for personal privacy. I also advocate for saving lives. And it is possible to do both. At Kinsa, **we do not share individual data, we share real-time population health insights.** Access to timely information is critical when it comes to public health, and the solutions essential for our country's welfare and health need to be advanced.*

What safeguards are in place to ensure data collected as part of the fight against COVID-19 are not sold to business partners or used for the development of other commercial products?

**Inder Singh Response:**

*There is a difference between selling personal information and selling aggregate population insights. There should be no problem with selling population health insights. Healthcare*

organizations already do this today -- including major institutions like Optum. If it is truly a population insight, there is no way to identify an individual. We should encourage business models to be developed that create societal benefits from population level insights. We must realize that **new and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy.**

Just as we currently incentivize the creation of drugs, diagnostics and vaccines, **we should be incentivizing the creation, adoption, integration and effective use of novel datasets by local, state and federal public health agencies.** And as I have said so often in my responses to questions posed during this hearing, we can and must also ensure that personal privacy protections also exist -- **We can have both: personal privacy and access to the novel datasets needed** to stop the spread of COVID-19. This is critical to the health and welfare of our fellow citizens and country at large. It is during crises like this that we realize health and public health is a most important essential element of our lives and economy.

To reiterate, population level insights like those Kinsa shares are not personal data, not deidentified data, not even metadata about a person. They are insights about the population. There is absolutely no way to identify an individual based on the percentage of people with fever or symptoms in a county. Kinsa's population health insights are available to the public and to public health first-responders at [Healthweather.us](https://www.healthweather.us). We believe this early warning system is critical infrastructure for our country to stop the next outbreak from becoming an epidemic.

7. Anonymization techniques are also critical for safeguarding consumers' privacy. Truly anonymized data can protect a consumer's personal information, like their geolocation, political opinions, or religious beliefs.

How do companies guarantee that every dataset they are storing contains truly anonymous data? And is the ability to re-identify data a part of the discussion in data-sharing arrangements?

### **Inder Singh Response:**

While there are legitimate reasons to share de-identified data, for example research purposes or to evaluate the impact of a particular program on participants, restrictions on attempts to re-identify the data should be incorporated into all agreements, and clauses should specifically state that individuals should not be re-identified, even by researchers.<sup>3</sup> The following policies should be encouraged in any agreements:

<sup>3</sup> We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.

1. *Limit sharing of de-identified data for purposes of research, program evaluation or other narrowly defined, legitimate purposes to provide communal value*
2. *Require any companies sharing de-identified data for the purposes above to include clauses that specifically state that individuals should not be re-identified, even by researchers, or combined with other data sets known to be able to re-identify data*

***Kinsa shares population health insights - the percent of people in a county who are ill.***  
*There is no way to identify an individual from this illness signal. To reiterate, such population level insights are not personal data, not deidentified data, not even metadata about a person. They are insights about the population. **We need not sacrifice personal privacy for the sake of public benefit. It is possible to successfully do both.***



## Sen. Blunt

As you know, this committee has prioritized drafting federal privacy legislation for the purpose of creating clear, baseline definitions and standards for data collection, storage, and use across industry sectors. Similarly, the bills before this committee attempt to create definitions to meet appropriate levels of consent and transparency for protecting consumers' privacy and security. In relation to COVID-19, the end users of specific data sets, like location data, are more likely to be governmental entities than commercial entities. Big data can be an incredible tool to better understand the spread of the virus, and the impact on communities across the country. Data can help identify resource deficits, inform governments and health care professionals to employ countermeasures at the appropriate time, and provide insight to the downstream economic effects of this pandemic.

However, U.S. commercial entities that would likely be collecting this data have very few guardrails on the collection and distribution of this data. Similarly, there are few requirements or regulations at federal and state levels which guide methodologies for anonymizing or pseudonymizing data. De-identifying data may result in greater data privacy and data security for consumers or individual citizens, but relies heavily on all of the entities involved in the collection and storage of that data making decisions based on best practices.

8. What efforts do you recommend that federal agencies undertake to ensure that data being used to track viral spread are upholding the highest possible standards for individual privacy and security?

## **Inder Singh Response:**

*Thank you, Senator Blunt. The tradeoff between personal information protection and providing information for society's benefits is a false one. **We can have both.** And we need both if we are going to successfully combat the second and third waves of COVID-19 coming which threaten the fabric of our healthcare system. Data collection, when done right, can create **enormous benefits for individuals and communities, while protecting individual privacy.** We must realize that **new and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy.** The following policies should be encouraged:*

- 1. Limit sharing of de-identified data for purposes of research, program evaluation or other narrowly defined, legitimate purposes to provide communal value*
- 2. Require any companies sharing de-identified data for the purposes above to include clauses that specifically state that individuals should not be re-identified, even by researchers, or combined with other data sets known to be able to re-identify data*

9. Does data lose any utility when it is de-identified or anonymized? Is it possible to have large data sets that are not tied to individual's identities, but which would still be useful for governments or public health-related end users?

**Inder Singh Response:**

*I believe there is a way to both protect personal privacy and also gather and share the information necessary to detect and effectively respond to outbreaks like COVID-19. Data collection, when done right, can create **enormous benefits for individuals and communities, while protecting individual privacy**. We have accomplished that at Kinsa. We do not share personal data, even if it is de-identified -- or anonymized -- data. With enough effort, even de-identified data has the potential to be re-identified. **We at Kinsa share population health insights** - the percent of people in a county who are ill. There is no way to identify an individual from this illness signal. While there are legitimate reasons for sharing de-identified data - for example, for research purposes or to evaluate the effectiveness of a health intervention - restrictions on attempts to re-identify the data, as I mentioned above, should be incorporated into all agreements.<sup>4</sup> There is no way to identify an individual from this system. We need not sacrifice personal privacy for the sake of public benefit. **We can have both.***

10. It is important to me that as government entities access commercially-collected or publicly available data, that those efforts are giving reasonable consideration to protecting individual privacy and security.

Are there any technologies that offer the opportunity to collect data that would be useful to a governmental pandemic response efforts, without resorting to surveillance methods that jeopardize individual privacy – like those which have been used recently by foreign governments?

**Inder Singh Response:**

*Senator, I agree that there are datasets coming out of the private sector that can add a tremendous amount of value for our government entities, and ultimately, help the American people. Just as we currently incentivize the creation of drugs, diagnostics and vaccines, **we should be incentivizing the creation, adoption, integration and effective use of novel datasets by local, state and federal public health agencies**. And we can and must also ensure that personal privacy protections also exist. We must also ensure that local public health departments have the necessary capabilities to make full use of data, and evaluate innovative technologies and approaches to improving public health. As it stands today, only a handful of federal agencies, and an even smaller number of teams within those federal agencies, are staffed, equipped, and funded enough to engage with new kinds of data or innovations. This risks that **many promising data sets and technologies will be overlooked**.*

*This has been our experience:*

<sup>4</sup> We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.

*I started Kinsa with a mission to curb the spread of infectious illness through earlier detection and earlier response, primarily because I didn't see others -- whether in government or in the private sector -- leveraging innovation to solve this problem. In particular, we sought to aggregate **new kinds of data**: data on where and when symptoms were starting, how fast they were spreading, and how long lasting or severe they got. This has enabled Kinsa to predict flu spread well before CDC, including predicting the severity of the 2017-2018 flu season (as [cited by the NY Times](#) in early 2018), which killed 80,000 Americans and resulted in tents outside of emergency rooms.*

*I believe that Kinsa's **population health insights** are incredibly valuable to the governmental pandemic response efforts, and do not jeopardize individual privacy. We do not share personal data; we share the percent of people in a county who are ill. Our population health insights are available to the public and to public health first-responders at [Healthweather.us](#). We believe this early warning system is **critical infrastructure for our country to stop the next outbreak from becoming an epidemic.***

## Sen. Cruz

11. Mr. Singh, since I joined the Senate my number one priority has been to ensure that Americans have access to jobs. Unfortunately as a result of state and local stay-at-home orders and cessation of business that is deemed non-essential, millions of Americans have lost their jobs and are struggling to find work. In March, the Federal Reserve estimated that the unemployment rate may eventually skyrocket to over 30%—a level that surpasses the Great Depression. It's clear we need to get people back to work once it is safe to do so, but figuring out when that is has been a struggle.

Now you mention in your testimony that one of your four key steps to stemming the spread of the virus is to implement antibody testing. I would agree with you, and in fact I recently sent a letter to Secretary Azar urging him to procure and distribute serological tests from the Strategic National Stockpile to help determine whether an individual has previously contracted and recovered from the coronavirus. Beyond helping stop the spread of the virus, this will also enable Americans who have antibodies to return to work and help get this economy rolling again.

Mr. Singh, what role do you see serological testing playing in helping get Americans back to work, and what can the Federal Government do to help the private sector manufacture and distribute serological tests?

### ***Inder Singh Response:***

*Thank you, Senator Cruz. Serological testing can play a very important role in helping get Americans back to work. The Federal Government can provide funding to companies working on these tests to help with manufacturing scale-up. With respect to distribution of the test, we believe that the first shipment of tests should go to essential workers starting with healthcare, police, emergency technicians, grocery store and pharmacy workers. Those with COVID-19 protective antibodies can safely return to the workforce without risk of being infected and without the need for PPE. This will allow those individuals to safely rejoin the workforce and provide essential care.*

*Senator Cruz, I also solicited input from a Kinsa board member, Dr. Beth Seidenberg, former Chief Medical Officer of Amgen, who suggests: The ideal test would be a point-of-care administered by a finger prick that would result in a blood sample being transferred to a test cassette that would provide a qualitative reading of the presence of SARS-CoV-2 developed IgM and/or IgG antibody. The test would have limited false positive or false negative interactions with other closely related viruses.*

12. A little over two weeks ago, the Johns Hopkins Center for Health Security published a report titled “*Modernizing and Expanding Outbreak Science to Support Better Decision*

*Making During Public Health Crises: Lessons for COVID-19 and Beyond.*” Although full of thought provoking ideas, one of the most notable was a recommendation to establish a “National Infectious Disease Forecasting Center,” similar to the National Weather Service. Much like the National Weather Service, this new infectious disease forecasting center would have both an operational role—providing the best modeling and forecasting to policy makers and public health professionals before, during, and after a disease outbreak—as well as a research role—providing a venue for academic, private sector, and governmental collaboration to improve models and encourage innovation.

What do you all think of this idea, and what do you all think the positives and negatives would be if such a concept was operationalized?

### ***Inder Singh Response:***

*We think the idea of having an independent entity specifically doing outbreak detection and monitoring is a good idea. The reason: in our experience, existing institutions like CDC do critical, life-saving work but have had difficulty adapting to new, 21<sup>st</sup> century technologies that enable early detection. They have also had difficulty engaging with the private sector. For those reasons, **I would encourage the creation of an independent entity** knowledgeable about 21<sup>st</sup> century technology and appropriately staffed to engage with the private sector where technology innovation is occurring.*

*As it stands today, only a handful of federal agencies, and an even smaller number of teams within those federal agencies, are staffed, equipped, and funded enough to engage with new kinds of data or innovations. This risks that **many promising data sets and technologies will be overlooked**. I would also encourage policies and funding for our local public health departments, so they may have the necessary capabilities to make full use of novel datasets, and may evaluate innovative technologies and approaches to improving public health. With more funding, local health departments can assist our federal efforts to stop the spread of illness -- they can hire the right experts and carry out the work to better integrate data and evaluate innovations.*

*I also believe there are technology solutions coming out of the private sector – such as Kinsa’s – that need to be accelerated to help stop the spread of COVID-19 and other outbreaks, but have of yet been able to capture the attention of the federal government.*

*This has been our experience:*

*I started Kinsa with a mission to curb the spread of infectious illness through earlier detection and earlier response, primarily because I didn’t see others -- whether in government or in the private sector -- leveraging innovation to solve this problem. In particular, we sought to aggregate **new kinds of data**: data on where and when symptoms were starting, how fast they were spreading, and how long lasting or severe they got. This has enabled Kinsa to predict flu spread well before CDC, including predicting the severity of the 2017-2018 flu season (as [cited](#)*

*by the NY Times in early 2018), which killed 80,000 Americans and resulted in tents outside of emergency rooms.*

*While we have continued to succeed in our predictions -- as exemplified by healthweather.us, a 2-week early warning system to first death from COVID-19 in 88% of states -- we continue to be overlooked by federal agencies, even after presenting peer-reviewed scientific literature supporting our work. I believe a key reason for this is that there is **only a very small group of people at the national level making critical decisions for the country**. This centralizes thinking, limits impact and reduces innovation. For the sake of the American people, we can do better.*

13. One of the big reasons weather forecasting works, if not the biggest, is how many observations—things like water temperature, barometric pressure, radio profiles of the atmosphere, etc.—are fed into the weather model. Now while collecting ocean temperatures from buoys, or pressure readings from weather balloons, doesn't really raise privacy concerns, collecting health observations almost certainly would.

How can we thread the needle—either in this concept or private sector modeling—of getting enough of the right kind of data to accurately model infectious disease outbreaks while still protecting the privacy and security of individuals?

#### ***Inder Singh Response:***

*I believe there is a way to both protect personal privacy and also gather and share the information necessary to detect and effectively respond to outbreaks like COVID-19. The tradeoff between personal information protection and providing information for society's benefits is a false one. **We can have both**. And we need both if we are going to successfully combat the second and third waves of COVID-19 coming which threaten the fabric of our healthcare system. Data collection, when done right, can **create enormous benefits for individuals and communities, while protecting individual privacy**. We have accomplished that at Kinsa. We do not share personal data, even if it is anonymized data. With enough effort, even anonymized data has the potential to be re-identified. We at Kinsa share **population health insights** - the percent of people in a county who are ill. There is no way to identify an individual from this illness signal. While there are legitimate reasons for sharing de-identified data - for example, for research purposes or to evaluate the effectiveness of a health intervention - restrictions on attempts to re-identify the data should be incorporated into all agreements.<sup>5</sup>*

<sup>5</sup> We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.

14. To date the State of Texas has reported thousands of cases of coronavirus, and hundreds of deaths related to complications from infection. To mitigate the risk of infection in Texas and across the country, the administration has restricted international travel, provided more access to medical supplies by involving the powers of the Defense Production Act, and cut red tape to expand access to testing. Congress also passed the CARES Act which provided \$377 billion in emergency loans for small businesses and directed \$100 billion to hospitals and healthcare providers. However, I believe much still needs to be done to finish this fight and recover once this is behind us.

In your expert opinions, what more needs to be done to beat this virus, and how can federal, state, and local governments work with private companies to both mitigate spread of the virus—both now and later this summer or fall—and recover quickly once the threat of this virus has passed?

### ***Inder Singh Response:***

*There are many systems that need to be put in place to ensure the appropriate readiness to enable recovery and help our country get back to work. Three gaps I'd like to highlight include:*

- 1. The necessity of an early detection and monitoring system*
- 2. The ability to contact trace for earlier intervention and isolation*
- 3. Antibody testing to determine immunity level of a population*

*With an **early detection and monitoring system**, we will understand (1) where outbreaks are cropping up, especially after we relax social distancing, allowing immediate action and interventions to stop the spread and (2) real-time knowledge of whether the chain of infection has been broken in the community thanks to the policy measures enacted. Using a method like Kinsa's, where **we share population health insights** - the percent of people in a county who are ill - would allow for an early warning system without sacrificing any personal privacy. Our population health insights are available to the public and to public health first-responders at [Healthweather.us](https://www.healthweather.us). We believe this early warning system is critical infrastructure for our country to stop the next outbreak from becoming an epidemic.*

***Contact tracing has been an effective way to get exposed citizens the tests and treatment they need, or encourage self-isolation to stop the spread of COVID-19.** I recently read a brilliant solution in [Kottke.org](https://www.kottke.org) that illustrated a contact tracing solution that would both protect personal privacy while also helping stop the spread of COVID-19. To explain briefly: contact tracing apps can work via Bluetooth and regularly send random unique "messages" of letters and numbers to any phones nearby. Because the messages are random and do not use GPS, they contain no personally-identifiable information. Any nearby phones with compatible contact tracing apps pick up and store the random messages sent by others in their vicinity. If any user tests positive for COVID-19, they or their doctor can input that information into the app. Any phones that hold a relevant "message" from the time the infected person would have been near them are notified that they were exposed. Throughout this process, no personally identifiable information has been captured, stored or shared. These are the types of technologies and*

*innovations our country needs, which both protect personal privacy while also giving us the information needed to stop the spread of illness and save lives.*

*Lastly, antibody testing would help us understand the level of COVID-19 immunity in the American population and help many of our fellow citizens feel safer about returning to work or venturing out to become reintegrated in society.*



**Sen. Fischer**

15. In terms of clusters of unusual fever that Kinsa thermometers detect, can you please quantify how strong the correlation has been with COVID-19 hotspots to date? Additionally, how are COVID-19-related fever statistics differentiated from individuals who have fever due to other illnesses?

***Inder Singh Response:***

*Thank you, Senator Fischer. Kinsa's atypical illness signal is highly correlated to COVID-19 outbreaks. On average, **we observe illness anomalies within states approximately 12-14 days before they report their first COVID-19 death.** For the 48 states in the continental US, Kinsa's atypical illness signal recorded illness anomalies at least five days prior to the first fatality in 42 states, or 88% of states. This means that Kinsa's atypical illness signal can serve as an early detection system in future outbreak monitoring.*

*Experts are still working on understanding and collecting the symptomology of COVID-19, and at this time, our system does not differentiate between fevers due to COVID-19, or other factors. But because of our scientific breakthrough in forecasting the flu 12+ weeks out, we are able to detect anomalous levels of fever, above and beyond what is appropriate for cold and flu. This unusual level of illness has been proven to be **highly correlated to COVID-19.** Think of Kinsa as a flashlight going off, illuminating areas where unusually high levels of illness are occurring and intervention is needed. I'd like to take a moment to emphasize that such a system is essential even outside of COVID-19 response. We cannot forget that even the seasonal flu can be very severe, and indeed, is severe in two of every 10 years. The 2017-2018 flu season claimed 80,000 U.S. lives and overflowed the healthcare system of the Southeastern United States. How do you stop an outbreak before it becomes an epidemic if you do not have knowledge of community spread in real time? The answer is you do not. You cannot. **An early warning system is necessary infrastructure for our country to stop the next epidemic.***

16. What obfuscation measures does Kinsa use to de-identify data and actively work to prevent unauthorized access to personally identifiable information?

***Inder Singh Response:***

*I believe there is a way to both protect personal privacy and also gather and share the information necessary to detect and effectively respond to outbreaks like COVID-19. The tradeoff between personal information protection and providing information for society's benefits is a false one. **We can have both.** And we need both if we are going to successfully combat the second and third waves of COVID-19 coming which threaten the fabric of our healthcare system. Data collection, when done right, can create*

enormous benefits for individuals and communities, while protecting individual privacy. We must realize that **new and novel data sets that help us do better analysis** -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy. We have accomplished that at Kinsa. **We share population health insights - the percent of people in a county who are ill.** There is no way to identify an individual from this illness signal. While there are legitimate reasons for sharing de-identified data - for example, for research purposes or to evaluate the effectiveness of a health intervention - restrictions on attempts to re-identify the data should be incorporated into all agreements.<sup>6</sup> This is because with enough effort, even de-identified -- or anonymized -- data has the potential to be re-identified. The following policies should be encouraged:

1. Limit sharing of de-identified data for purposes of research, program evaluation or other narrowly defined, legitimate purposes to provide communal value
2. Require any companies sharing de-identified data for the purposes above to include clauses that specifically state that individuals should not be re-identified, even by researchers, or combined with other data sets known to be able to re-identify data

To reiterate, Kinsa's population level insights are not personal data, not deidentified data, not even metadata about a person. They are insights about the population. There is absolutely no way to identify an individual based on the percentage of people with fever or symptoms in a county. Our population health insights are available to the public and to public health first-responders at [Healthweather.us](https://www.healthweather.us). We believe this early warning system is critical infrastructure for our country to stop the next outbreak from becoming an epidemic. There is no way to identify an individual from this system. **We need not sacrifice personal privacy for the sake of public benefit. It is possible to successfully do both.**

<sup>6</sup> We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.

## Sen. Moran

17. Many of the discussed proposals related to utilizing “big data” to fight against the spread against coronavirus rely upon the concepts of anonymized and aggregated data to protect the personal identity of individuals that this information pertains to and prevent consumer harms that could result. As such, many members on this Committee have spent significant time and energy drafting federal privacy legislation that tries to account for practices such as these that prevent harmful intrusions into consumers’ privacy while also preserving innovative processing practices that could utilize such information responsibly without posing risks. *That being said, do the witnesses have any policy recommendations for the Committee as it relates to effectively defining technical criteria for “aggregated” and “anonymized” data, such as requiring companies to publicly commit that they will refrain from attempting to re-identify data to a specific individual while adopting controls to prevent such efforts?*

## **Inder Singh Response:**

*Thank you, Senator Moran. I believe the tradeoff between personal information protection and providing information for society’s benefits is a false one. **We can have both.** Data collection, when done right, can create enormous benefits for individuals and communities, while protecting individual privacy. We must realize that **new and novel data sets that help us do better analysis** -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy. We have accomplished that at Kinsa. **We share population health insights - the percent of people in a county who are ill.** There is no way to identify an individual from this illness signal. While there are legitimate reasons for sharing de-identified data - for example, for research purposes or to evaluate the effectiveness of a health intervention - restrictions on attempts to re-identify the data should be incorporated into all agreements.<sup>7</sup> This is because with enough effort, even de-identified -- or anonymized -- data has the potential to be re-identified. The following policies should be encouraged:*

- 1. Limit sharing of de-identified data for purposes of research, program evaluation or other narrowly defined, legitimate purposes to provide communal value*
- 2. Require any companies sharing de-identified data for the purposes above to include clauses that specifically state that individuals should not be re-identified, even by researchers, or combined with other data sets known to be able to re-identify data*

18. Consumer data has tremendous benefits to society, as is clearly evident in the fight against the COVID-19 outbreak. Big data and the digitized processes and algorithms

<sup>7</sup> *We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.*

that technology companies are developing have led to an entirely new sector of the global economy. *Are you satisfied that the technology industry is striking an appropriate balance between producing services that better our ability to solve problems, as is clear in the fight against COVID-19, versus their production of products that increase their bottom line and generate profit? Are you satisfied that the United States government is striking an appropriate balance between supporting these companies in addressing COVID-19 versus ensuring we conduct adequate oversight of the industries' activities?*

**Inder Singh Response:**

*Big tech has had its fair share of battles -- many of them quite warranted -- over privacy and the collection of personal data. With the white-hot spotlight on the industry as we all try to determine the best way to balance personal freedoms with public health, what better time for the industry to make a hard stance in favor of protecting personal health data? We have found a way to do this at Kinsa and recommend every company working in this space do the same. We do not share personal data; **We share population health insights** - the percent of people in a county who are ill. There is no way to identify an individual from this illness signal.*

*In a recent op-ed in [InsideHealthPolicy](#) Dr. Nirav Shah, former State Health Commissioner for the state of New York, stated that “[d]ata from private companies like Unacast and Kinsa could work for short-term decision-making in the health care community, but also eventually be built into a bigger national surveillance system...This is repurposing existing data for the public good. Private companies doing things for the public good.” These are but 2 examples in a sea of several where private sector innovation can drive value for the healthcare system -- in this case, by helping to stop the spread of a pandemic that is costing American lives daily -- and for our country, while still respecting individual privacy.*

*In our experience, existing institutions like the CDC do critical, life-saving work but have had difficulty adapting to new, 21<sup>st</sup> century technologies that enable early detection. They have also had difficulty engaging with the private sector. **I would encourage the creation of an arm of the CDC** appropriately staffed to engage with the private sector and startups in particular, where technology innovation is occurring fastest. As it stands today, only a handful of federal agencies, and an even smaller number of teams within those federal agencies, are staffed, equipped, and funded enough to engage with new kinds of data or innovations. This risks that **many promising data sets and technologies will be overlooked**. I would also encourage policies and funding for our local public health departments, so they may have the necessary capabilities to make full use of novel datasets, and may evaluate innovative technologies and approaches to improving public health. With more funding, local health departments can assist our federal efforts to stop the spread of illness -- they can hire the right experts and carry out the work to better integrate data and evaluate innovations.*

19. Consumer trust is essential to both the United States government and to the companies whose products we use every day. We need to work to maintain that trust and ensuring that the big data being used to analyze the COVID-19 outbreak was collected and processed in a manner that aligns with our principles is important to my constituents. *How can we adequately ensure that the data being used to address COVID-19 is sourced and processed in a manner that ensures consumer trust is not being violated, while allowing the innovation and success we've seen continue to grow?*

**Inder Singh Response:**

*I agree that consumer trust is essential, and that personally-identifiable information should never be used without the user's explicit permission and ongoing awareness. But with proper user consent, insights drawn from aggregated, population data should be accessible for the public good. Data collection, when done right, **can create enormous benefits for individuals and communities, while protecting individual privacy.** We too often focus on the collection of personal data, rather than on building protocols for the protection of personal data; we focus on the private benefits that accrue to companies that monetize data but often do not consider the possibilities of data use associated with public benefit; we focus on the right to privacy, but in doing so, we ignore another important right – **the right to information.** To be clear, I advocate for personal privacy. I also advocate for saving lives. And it is possible to do both. At Kinsa, **we do not share individual data, we share real-time population health insights.** Access to timely information is critical when it comes to public health, and the solutions essential for our country's welfare and health need to be advanced.*

20. It is important to remember that the internet is a global network and that no matter how secure we make our networks, they remain vulnerable to bad actors, corruption, and misguided influence from around the world. *Can you comment on the practices we've seen used by companies and international partners to ensure the data used to address COVID-19 is both accurately sourced and stored in a manner that is secure?*

**Sen. Blackburn**

21. How do you see HIPAA interacting with your worldview of the tech industry?

***Inder Singh Response:***

*Thank you, Senator Blackburn. HIPAA's intentions are sound: to protect personal health information. While provisions within it are relevant, HIPAA is restricted to covered entities such as health insurers and health providers. The real question here is how to protect personal information, including personal health information, while enabling technological innovation that creates societal benefits, including population level insights necessary to detect outbreaks, predict impact and ensure individual are cared for early (e.g., contact tracing solutions). The tradeoff between personal information protection and providing information for society's benefits is a false one. **I believe we can have both.***

*Data collection, when done right, can create **enormous benefits for individuals and communities, while protecting individual privacy.** We must realize that **new and novel data sets that help us do better analysis** -- such as predicting where and when outbreaks like COVID-19 will spread -- **are of immense value to our public health system and our economy.** We have accomplished that at Kinsa. **We share population health insights** - the percent of people in a county who are ill. There is no way to identify an individual from this illness signal.*

*Contact tracing can also be done in a way that protects individual privacy while helping to stop the spread of COVID-19. I recently read a [brilliant comic in Kottke.org](#) that illustrated a contact tracing solution that would both protect personal privacy while also helping stop the spread of COVID-19. To explain briefly: contact tracing apps can work via Bluetooth and regularly send random unique "messages" of letters and numbers to any phones nearby. Because the messages are random and do not use GPS, they contain no personally-identifiable information. Any nearby phones with compatible contact tracing apps pick up and store the random messages sent by others in their vicinity. If any user tests positive for COVID-19, they or their doctor can input that information into the app. Any phones that hold a relevant "message" from the time the infected person would have been near them are notified that they were exposed. Throughout this process, no personally identifiable information has been captured, stored or shared. These are the types of technologies and innovations our country needs, which both protect personal privacy while also giving us the information needed to stop the spread of illness and save lives.*

22. How do you envision working with the CDC to develop the updated surveillance system (which was given \$500 million in the recently passed CARES Act) while protecting health information and thereby allow CDC to use their expertise – epidemiology that inherently seeks to protect health information – with big tech's powerful data collection and analysis tools?



### **Inder Singh Response:**

*We would welcome the opportunity to work with the CDC. We respect and admire the CDC, and both the institution, and the American people, would benefit from closer collaboration between the private sector and the CDC.*

*I believe the CDC needs to be equipped with the **right people, processes and incentives to (a) explore and embrace the use of modern 21st century technologies,** (b) engage with the private sector and particularly with startups making the biggest strides in the areas of healthcare and public health. We as a country cannot be constrained by traditional thinking in public health. As is evidenced by the COVID-19 epidemic, we must use both traditional thinking & activities as well as novel, technologically advanced approaches to stop the spread of illness.*

*I started Kinsa with a mission to curb the spread of infectious illness through earlier detection and earlier response, primarily because I didn't see others -- whether in government or in the private sector -- leveraging innovation to solve this problem. In particular we sought to aggregate new kinds of data: data on where and when symptoms were starting, how fast they were spreading, and how long lasting or severe they got. This is what has enabled Kinsa to predict flu spread well before CDC, including predicting severity of the 2017-2018 flu season -- which killed 80,000 Americans and resulted in tents outside of Emergency Rooms -- months before CDC (as cited by the NY Times in early 2018). While we have succeeded in building these data sets, and making use of the population insights from them -- as exemplified by healthweather.us which has provided a two week early warning system to first death from COVID-19 in nearly every location in which we hotspot unusual illness spread -- we were overlooked many times by federal agencies, even after presenting peer-reviewed scientific literature supporting our work.*

*The implications of a system like Kinsa's -- which is low-cost, participatory, resilient and absolutely protects the privacy of an individual while democratizing information -- is necessary for communities, our healthcare system, and our country to respond to outbreaks. There are other private sector companies and startups that are doing similarly critical public health work -- we must realize that **new and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy.***

*This leads me to a conclusion: Just as we currently incentivize the creation of drugs, diagnostics and vaccines, **we should be incentivizing the creation, adoption, integration and effective use of novel datasets by local, state and federal public health agencies.** And as I have said so often in my responses to questions posed during this hearing, we can and must also ensure that personal privacy protections also exist -- **We can have both: personal privacy and access to the novel datasets needed** to stop the spread of COVID-19. This is critical to the health and welfare of our fellow citizens and country at large. It is during crises like this that we realize health and public health is a most important essential element of our lives and economy.*

23. Today we are giving into state surveillance for the sake of saving thousands of lives that might otherwise be lost to coronavirus. The CDC is already relying on data analytics from mobile ad providers to track the spread of the disease. How can we ensure the data collection will only be done for the limited purposes of the emergency, with safeguards to ensure anonymity? On retention time, when should the data be deleted? Who has the right to that deletion – the federal government or the individuals themselves? Most importantly, what duty do tech companies owe to protect consumer privacy, even during a global pandemic?

**Inder Singh Response:**

*At Kinsa, we believe personally-identifiable information should never be used without the user's explicit permission and ongoing awareness. But with proper user consent, insights drawn from aggregated, anonymized data should be accessible for the public good. Data collection, when done right, **can create enormous benefits for individuals and communities, while protecting individual privacy.** We too often focus on the collection of personal data, rather than on building protocols for the protection of personal data; we focus on the private benefits that accrue to companies that monetize data but often do not consider the possibilities of data use associated with public benefit; we focus on the right to privacy, but in doing so, we ignore another important right – **the right to information.** To be clear, I advocate for personal privacy. I also advocate for saving lives. And it is possible to do both. At Kinsa, **we do not share individual data, we share real-time population health insights - the percent of people in a county who are ill.** There is no way to identify an individual from this illness signal. While there are legitimate reasons for sharing de-identified data - for example, for research purposes or to evaluate the effectiveness of a health intervention - restrictions on attempts to re-identify the data should be incorporated into all agreements.<sup>8</sup> Access to timely information is critical when it comes to public health, and the solutions essential for our country's welfare and health need to be advanced.*

24. Foreign countries like South Korea, Taiwan, Singapore, and Israel swiftly mobilized collection of cell phone location data to track the spread of the virus and map out infection hot zones. Israel just released an app that allows the public to track whether they have may visited a location that put them into contact with an infected individual. Is it even possible to adopt similar measures while still balancing protections for privacy and civil liberties?

<sup>8</sup> *We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.*



### **Inder Singh Response:**

Yes, I believe it is possible. The tradeoff between personal information protection and providing information for society's benefits is a false one. **We can have both.** And we need both if we are going to successfully combat the second and third waves of COVID-19 coming which threaten the fabric of our healthcare system. **Data collection, when done right, can create enormous benefits for individuals and communities, while protecting individual privacy.** We have accomplished that at Kinsa. We do not share personal data, even if it is anonymized data. With enough effort, even anonymized data has the potential to be re-identified. **We share population health insights** - the percent of people in a county who are ill. There is no way to identify an individual from this illness signal.

It is even possible to do contact tracing in a way that protects individual privacy while helping to stop the spread of COVID-19. I recently read a [brilliant comic in Kottke.org](http://brilliant.comic.in.kottke.org) that illustrated a contact tracing solution that would both protect personal privacy while also helping stop the spread of COVID-19. To explain briefly: contact tracing apps can work via Bluetooth and send a series of random "messages" of letters and numbers to any phones nearby. Because the messages are random and do not use GPS, they contain no personally-identifiable information. Any nearby phones with compatible contact tracing apps pick up and store random messages sent by others in their vicinity. If any user tests positive for COVID-19, they or their doctor can input that information into the app. Any phones that hold a relevant "message" from the time the infected person would have been near them are notified that they were exposed. Throughout this process, no personally identifiable information has been captured, stored or shared.

**Sen. Lee**

25. To date, what specific data (or types of data) are companies (or your company) currently collecting for COVID-19 related purposes? What specific data (or types of data) are governments and health officials seeking for COVID-19 related purposes?

**Inder Singh Response:**

*Thank you, Senator Lee. Kinsa aggregates temperature and symptom inputs that we share as population health insights. Kinsa has been using these **population health insights** to track and predict flu for the past 6 years. Until January 2020, we were able to predict the spread of flu-like illness 3 weeks in advance on a county by county basis. In early February, we had a scientific breakthrough that allowed us to forecast flu 12-14 weeks out. With this breakthrough, we were able to subtract expected cold and flu numbers from actual observed cold/flu to detect unusual fever clusters - "hotspots" - indicating community spread of COVID-19. We made these population health insights available to the public and to public health first-responders at [Healthweather.us](https://www.healthweather.us). We believe this early warning system is critical infrastructure for our country to stop the next outbreak from becoming an epidemic. There is no way to identify an individual from this system. **We need not sacrifice personal privacy for the sake of public benefit. It is possible to successfully do both.***

26. Most tech companies currently claim that the data being gathered is being "anonymized" so that a specific person is not identifiable.

What specific steps are companies (or your company) taking to anonymize this data?

**Inder Singh Response:**

*Data collection, when done right, can create **enormous benefits for individuals and communities, while protecting individual privacy**. We must realize that **new and novel data sets that help us do better analysis** -- such as predicting where and when outbreaks like COVID-19 will spread -- **are of immense value to our public health system and our economy**. The following policies should be encouraged:*

- 1. Limit sharing of de-identified data for purposes of research, program evaluation or other narrowly defined, legitimate purposes to provide communal value*
- 2. Require any companies sharing de-identified data for the purposes above to include clauses that specifically state that individuals should not be re-identified, even by researchers, or combined with other data sets known to be able to re-identify data*

***At Kinsa, we do not share individual data, we share real-time population health insights. There is no way to identify an individual from this illness signal.<sup>9</sup> While there are legitimate reasons for sharing de-identified data - for example, for research purposes or to evaluate the effectiveness of a health intervention - restrictions on attempts to re-identify the data should be incorporated into all agreements.***

Certain data may not necessarily be considered personally identifiable, but with enough data points, you could identify a specific person. How can we ensure that data is truly anonymous and is not traceable back to an individual person?

Can effective contact tracing be conducted with “anonymized data”? Or will it require personally identifiable information?

***Inder Singh Response:***

*There are ways to use location based tracking and preserve personal privacy. **It is a false tradeoff – you can have both: personal privacy and access to the information needed to contact trace to stop the spread of COVID-19.** I recently read a [brilliant comic in Kottke.org](#) that illustrated a contact tracing solution that would both protect personal privacy while also helping stop the spread of COVID-19. To explain briefly: contact tracing apps can work via Bluetooth and send a series of random “messages” of letters and numbers to any phones nearby. Because the messages are random and do not use GPS, they contain no personally-identifiable information. Any nearby phones with compatible contact tracing apps pick up and store random messages sent by others in their vicinity. If any user tests positive for COVID-19, they or their doctor can input that information into the app. Any phones that hold a relevant “message” from the time the infected person would have been near them are notified that they were exposed. Throughout this process, no personally identifiable information has been captured, stored or shared.*

27. Since the beginning of this COVID-19 crisis, has a federal agency, a state government, or local government requested a company or association to gather any specific consumer data?

To your knowledge, are there any current COVID-19 related data sharing agreements in place between governments and private sector organizations?

<sup>9</sup> *We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.*

To your knowledge, has any federal, state, or local law enforcement used private sector collected data to enforce any COVID-19 related government orders or requirements?

***Inder Singh Response:***

*Not to our knowledge. As it stands today, only a handful of federal agencies, and an even smaller number of teams within those federal agencies, are staffed, equipped, and funded enough to engage with new kinds of data or innovations. This risks that **many promising data sets and technologies will be overlooked**. I would also encourage policies and funding for our local public health departments, so they may have the necessary capabilities to make full use of novel datasets, and may evaluate innovative technologies and approaches to improving public health. With more funding, local health departments can assist our federal efforts to stop the spread of illness -- they can hire the right experts and carry out the work to better integrate data and evaluate innovations.*

*I also believe there are technology solutions coming out of the private sector – such as Kinsa’s – that need to be accelerated to help stop the spread of COVID-19 and other outbreaks, but have of yet been able to capture the attention of the federal government.*

*This has been our experience:*

*I started Kinsa with a mission to curb the spread of infectious illness through earlier detection and earlier response, primarily because I didn’t see others -- whether in government or in the private sector -- leveraging innovation to solve this problem. In particular, we sought to aggregate **new kinds of data**: data on where and when symptoms were starting, how fast they were spreading, and how long lasting or severe they got. This has enabled Kinsa to predict flu spread well before CDC, including predicting the severity of the 2017-2018 flu season (as [cited by the NY Times](#) in early 2018), which killed 80,000 Americans and resulted in tents outside of emergency rooms.*

*While we have continued to succeed in our predictions -- as exemplified by [healthweather.us](#), a 2-week early warning system to first death from COVID-19 in 88% of states -- we continue to be overlooked by federal agencies, even after presenting peer-reviewed scientific literature supporting our work. I believe a key reason for this is that there is **only a very small group of people at the national level making critical decisions for the country**. This centralizes thinking, limits impact and reduces innovation. For the sake of the American people, we can do better.*

**Sen. Young**

28. Mr. Singh, what datasets can the federal government make available that would assist in using data analytics to respond to coronavirus and once the pandemic has ended, how do you expect day to day life to change based on the unintended lessons we've learned during this time of social distancing?

***Inder Singh Response:***

*Thank you, Senator Young. The tradeoff between personal information protection and providing information for society's benefits is a false one. **We can have both.** And I hope that after this pandemic has ended, more people have the realization that we can have both, and that we can create innovations that allow for both.*

*It is critical that we have an early warning system for spreading illness, and not just for COVID-19 but for any number of contagious illnesses that unnecessarily claim American lives. As a reminder, the 2017-2018 flu season claimed 80,000 U.S. lives and overflowed the healthcare system of the Southeastern United States. We forget that even the seasonal flu can be very severe, and indeed, is severe in two of every 10 years. How do you stop an outbreak before it becomes an epidemic if you do not have knowledge of community spread in real time. The answer is you do not. You cannot. **An early warning system is necessary infrastructure for our country.***

## Sen. Scott

29. For months, Communist China lied about the Coronavirus data, the spread of the virus, and their response. They silenced critics and those trying to alert the Chinese people to this public health crisis. The lack of usable data coming out of Communist China cost lives and put the world behind on response efforts, including here in the United States.

As we work to keep American families healthy, how can we follow the lead of countries with low case counts, like South Korea, using technology and data collection, without infringing on our citizens' rights and privacy?

### **Inder Singh Response:**

*Thank you, Senator Scott. South Korea's massive testing and data collection effort was a version of an early warning system born out of necessity to stop COVID-19. How do you stop an outbreak before it becomes an epidemic if you do not have knowledge of community spread in real time. The answer is you do not. You cannot. **An early warning system is necessary infrastructure for our country.** Data collection, when done right, can create enormous benefits for individuals and communities, while protecting individual privacy. **We can have both.** And we need both if we are going to successfully combat the second and third waves of COVID-19 coming which threaten the fabric of our healthcare system. We have accomplished that at Kinsa. We do not share personal data, even if it is de-identified -- or anonymized -- data. With enough effort, even de-identified data has the potential to be re-identified. **We share population health insights** - the percent of people in a county who are ill. There is no way to identify an individual from this illness signal. While there are legitimate reasons for sharing de-identified data - for example, for research purposes or to evaluate the effectiveness of a health intervention - restrictions on attempts to re-identify the data should be incorporated into all agreements.<sup>10</sup> We need not sacrifice personal privacy for the sake of public benefit.*

<sup>10</sup> *We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.*

## Ranking Member Cantwell

30. Mr. Singh, I applaud you for developing a valuable and innovative service that seeks to protect public health while complying with privacy regulations – as you make clear in your privacy policy, not just on the Federal level, but also on a state-by-state basis as well. Throughout this crisis, we have frequently found ourselves behind the curve and I can't help thinking how badly we need better methods for detecting early signals, such as the system you and your team at Kinsa have developed.

I'd like to better understand how your system works, what it can tell us, and what you have been able to learn. In New Orleans in late February, coronavirus was undetected in the population, and the city proceeded with Mardi Gras celebrations. In retrospect, we now know that Mardi Gras was probably a super-spreader event. Based on data collected by Kinsa from New Orleans and surrounding parishes from mid-February through the present, could you summarize what you were able to learn from this outbreak? When did you first detect signs of the arrival and dispersal of COVID-19 in the population? What early insights were you able to gain? Were there early warnings that went undetected? What could have been done better, either by Kinsa or by the city of New Orleans? What lessons have you learned that could help all of us be better prepared for the sudden arrival of a pandemic?

### ***Inder Singh Response:***

*Thank you, Ranking Member Cantwell. In the case of COVID-19, our atypical illness map started monitoring for unusual fever clusters nationwide on March 1. In Orleans Parish, we first flagged "atypical" illness levels (higher than expected levels of fever in the population) on **March 6<sup>th</sup>**, and saw atypical illness levels in nearby parishes as early as **March 4<sup>th</sup>**. I will admit that our atypical illness signal did not detect New Orleans to be as much of a hot-spot that we know it to be today. Our hypothesis is that many people came into town for Mardi Gras, and that community spread occurred very rapidly. Those who came were not local, and ended up spreading COVID-19 back in their hometowns.*

*We will certainly benefit from more thermometer coverage in the New Orleans area, and from closer partnership with city/metro governments and public health organizations. Unfortunately at the time of Mardi Gras, almost no local governments were taking meaningful actions. In the context of an epidemic, or even a threat of an epidemic, a large event like Mardi Gras absolutely should not be allowed.*

*I believe four key steps are necessary to stem the growth of an epidemic:*

- 1) An early warning system to understand where and when illness is spreading*
- 2) Widespread testing capabilities*
- 3) Treatment and isolation of those infected*
- 4) Antibody testing to determine the immunity level of a population*

*In the absence of widespread testing, it is even more critical that we have an early warning system and the means to understand where outbreaks might be occurring. Kinsa is an early warning system. Think of it as a flashlight going off, illuminating a geography and saying, “send the test kits in, because something unusual is happening.” This **real-time information** on where and when illness is starting, and where community spread is occurring, is vital in appropriately allocating limited supplies and manpower to the areas most in need of early intervention.*

31. Science and technology will be critical drivers of our response to COVID-19, and we have seen many examples of data being used in positive ways – from the University of Washington’s forecasts of hospital needs to Johns Hopkins’ maps of disease spread. These are leading examples of how firms can innovate while protecting other equities, like privacy. What recommendations do you have to encourage further innovation to fight the virus? How do we encourage technologists to help people transition to regular life while preparing for future pandemic incidents? What are the best practices you have seen in innovating in the fight against COVID-19 that support privacy rights?

***Inder Singh Response:***

*Senator, I agree that data can be used in positive ways to help with our COVID-19 response, and strongly believe our country **should encourage** business models to be developed that create societal benefits from population health insights. Indulge me for a moment while I build to a point:*

*First: A good friend and brilliant innovator, Andy Palmer, CEO of Tamr, and one of the most successful angel investors in the country, often says that in today’s world, data is a currency. If we look at data that way for a moment, there are a few things to be realized -- but first and foremost, **data is valuable.***

*Second: Analysis is only as good as the data inputs. If the data inputs are bad, the resulting analysis and data output is bad. As goes the old saying famous amongst MIT graduates and anyone who works in data analysis: “garbage in equals garbage out.”*

*Given this, we must realize that **new and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy.***

*So that leads me to a conclusion: Just as we currently incentivize the creation of drugs, diagnostics and vaccines, **we should be incentivizing the creation, adoption, integration and effective use of novel datasets by local, state and federal public health agencies.** And as I have said so often in my responses to questions posed during this hearing, we can and must also ensure that personal privacy protections also exist -- **We can have both: personal privacy and access to the novel datasets needed** to stop the spread of COVID-19. This is critical to the health and welfare of our fellow citizens and country at large. It is during crises like*



*this that we realize health and public health is a most important essential element of our lives and economy.*

*One way we can strengthen our systems is by ensuring that local public health departments -- and not just well funded federal agencies -- have the necessary capabilities to make full use of data, and evaluate innovative technologies and approaches to improving public health. As it stands today, only a handful of federal agencies, and an even smaller number of teams within those federal agencies, staffed, equipped, and well-funded enough to engage with new kinds of data or innovations. This risks that **many promising data sets and technologies will be overlooked**. We must create departments in federal and local government that can hire people with the necessary expertise and carry out the work to better integrate data and evaluate innovations. This will in turn lead to more rapid progress in the use of 21st century technologies for improving public health.*

32. Frequently, data used to combat COVID-19 is described as “anonymized” or “aggregated” or “de-identified,” and these terms are meant to convey that data will be used or shared in a privacy-protective manner.

How do you define “anonymized,” “aggregated,” and “de-identified” data? What are the best practices to ensure that the data remains anonymous?

***Inder Singh Response:***

*De-identified data is stripped of identifiers like email address, phone number, name, etc. However, in my opinion, de-identified data risks being re-identified. With enough effort, most de-identified datasets can be identified. I believe de-identified data should only be provided when necessary, for example, to health researchers or to evaluate the effectiveness of health programs. But any contracts should specifically outline how the data cannot be combined with other data sets in order to re-identify an individual.*

*Anonymized data could mean de-identified, or it could mean fully anonymized in some other way - like through aggregation - so that no retracing can happen.*

*Aggregated data can mean several things, and usually, is a combination of data points resulting in a sum or a total. Typically, is not identifiable to an individual . At Kinsa, fever data is aggregated to do population level analysis and what we share is the percent of people in a county who are ill.*

*Data collection, when done right, **can create enormous benefits for individuals and communities, while protecting individual privacy**. We have accomplished that at Kinsa. We do not share personal data, even if it is de-identified -- or anonymized -- data. With enough effort, even anonymized data has the potential to be re-identified. **We share population health***

*insights* - the percent of people in a county who are ill. There is no way to identify an individual from this illness signal. While there are legitimate reasons for sharing de-identified data - for example, for research purposes or to evaluate the effectiveness of a health intervention - restrictions on attempts to re-identify the data should be incorporated into all agreements.<sup>11</sup> We need not sacrifice personal privacy for the sake of public benefit. The tradeoff between personal information protection and providing information for society's benefits is a false one. **We can have both.** And we need both if we are going to successfully combat the second and third waves of COVID-19 coming which threaten the fabric of our healthcare system.

<sup>11</sup> We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.

## Sen. Blumenthal

Kinsa has marketed its aggregate datasets as an indicator of community wellness and suggested to public health officials to use its map to “identify areas where illness levels are unusually high, and investigate” and to “gauge whether measures taken are working to slow the spread” of COVID-19.

However, significant regions of the United States are not covered by Kinsa’s COVID-19 map due to a lack of data, including states that are anticipated to be future epicenters of the pandemic. Moreover, Kinsa’s thermometers currently cost \$36 and \$70 on its site, and are unavailable for purchase, putting them far outside the reach of many households, particularly those affected by new financial pressures and less technologically savvy audiences.

We have seen countless examples where missing data and unaccounted variables in big data analytics leads to incorrect analysis that marginalizes or harms vulnerable populations. This is of pressing concern if Kinsa’s data may be used to make decisions of the allocation of public health resources. Simply put, if datasets are leading public health officials to overlook less well-off communities, then we could be creating new public health disasters.

33. How does Kinsa identify, account for, and address the exclusion of particular demographics from its aggregate health data, and what information about these factors and methods is provided to public health officials to ensure that inherent sampling biases do not lead to faulty decision-making?

## **Inder Singh Response:**

*I agree that vulnerable populations, particularly those being disproportionately affected by spreading illness, need to be included in any data sets, and this is why Kinsa has undertaken initiatives like our school health program, FLUency, specifically to put our product in the hands of the marginalized. By securing corporate sponsorships to cover the costs of this program, **Kinsa has given away 200,000 Kinsa smart thermometers to primarily Title 1 schools in under-resourced communities, balancing out retail sales for those who can and do pay for innovative health technology tools. I agree that an optimized early warning system should represent all populations and demographics, but **we must start somewhere.*****

*To ensure that our sample population will be effective in uncovering hotspots, we especially focus on three key audiences:*

- 1. Larger households, where illnesses such as COVID-19 are likely to linger longer, leading to unwitting spread – in the case of COVID-19, once social distancing regulations are relaxed*
- 2. Families in underserved communities, which typically seek care and treatment later, potentially spreading the virus before testing and treatment occur. Free*

*thermometers donated to families in Title 1 schools through our school health program, FLUency, make up 20% of our existing thermometers in use today. We are expanding this program nearly 3X this coming Fall, to reach nearly 4,000 underserved elementary schools across the country.*

*3. Families of first responders who are most likely to come in contact with spreading contagious illnesses such as COVID-19*

*Furthermore, for every thermometer purchased through our website, Kinsa is donating 1 QuickCare thermometer to a family in need. We are also endeavoring to work with state and local governments so that they may purchase and distribute Kinsa thermometers to communities in need within their states.*

## Sen. Schatz

34. Companies' datasets have been used to create models to forecast the spread of the pandemic. However, according to a recent Pew study, only 80% of Americans have access to a smartphone. A lot fewer use smart thermometers.

This Committee is well acquainted with the digital divide and the discriminatory impact caused by the lack of availability and access to broadband and smart technologies. Accordingly, can you assure the Committee that the datasets of your companies or member companies' are truly non-biased representations of the population, and will you commit to have these datasets audited by independent experts to ensure we are not making critical-decisions regarding the pandemic based on biased data?

### *Inder Singh Response:*

*Thank you, Senator Schatz. We agree that vulnerable populations, particularly those being disproportionately affected, need to be included in any data sets that analyze the spread of disease. However, disease spread typically occurs community-wide, and does not adhere to race, income, smartphone status, or geographic bounds, and community transmission affects adjacent communities. As evidenced by COVID-19, this can happen within days. Thus, we first and foremost need to at least understand how fast a disease is spreading in a geography, and then we must always endeavor to understand how fast it is spreading within subcommunities. An optimized early warning system would do both. **We must start somewhere** if we are going to achieve the optimal state.*

*Kinsa's intention has always been to create an early warning system for spreading illness. To accomplish this, we target three audiences:*

- 1. Larger households, where illnesses such as COVID-19 are likely to linger longer, leading to unwitting spread – in the case of COVID-19, once social distancing regulations are relaxed*
- 2. Families in underserved communities, which typically seek care and treatment later, potentially spreading the virus before testing and treatment occur. Free thermometers donated to families in Title 1 schools through our school health program, FLUency, make up 20% of our existing thermometers in use today. We are expanding this program this coming Fall. This program is especially important, given that the same Pew study which cited 80% of Americans having access to smartphones also provided the breakdown showing that **individuals likely to have elementary-aged children are more highly represented** - 96% in ages 18-29, and 92% in ages 30-49.*
- 3. Families of first responders who are most likely to come in contact with spreading contagious illnesses such as COVID-19*

Furthermore, and since near the beginning of the COVID-19 crisis, for every thermometer purchased through our website, Kinsa is donating 1 QuickCare thermometer to a family in need. We are also endeavoring to work with state and local governments so that they may purchase and distribute Kinsa thermometers to communities in need within their states. An early warning system, which is what Kinsa is building, does not require 100% of the population to use the product. The reason we specifically focus on distributing our product to families is because larger households are often spreaders of illness. A seminal study conducted by the [University of Utah](#) and published in the [Journal of Clinical Infectious Diseases](#) found that a family with 6 children has a virus in their home 45 weeks of the year; a household with no children has a virus in their home a mere 3-4 weeks of the year. As long as Kinsa thermometers are distributed across the country, it is possible to provide tremendous value for our nation and to save thousands of lives without 100% of Americans participating.

## Sen. Peters

35. The one thing that has been absent from this discussion is that neither the federal government nor the private sector have adequately anticipated nor met the demands for personal protective equipment. Even basic things like masks and gloves have been inaccessible. Our nation has unparalleled resources in the supply chain and manufacturing space.

From a data perspective—where have failures been and what improvements do you recommend?

### **Inder Singh Response:**

*Thank you, Senator Peters. **An early warning system is necessary infrastructure for our country.** How do you stop an outbreak before it becomes an epidemic if you do not have knowledge of community spread in real time? The answer is you do not. You cannot. This has been abundantly clear with our challenges in procuring adequate personal protective equipment supply. Had an early warning system been in place, we would have known where to send our limited stock of PPE and of test kits. This real-time information on where and when illness is starting, and where community spread is occurring, is vital in appropriately allocating limited supplies and manpower to the areas most in need of early intervention.*

*We must realize that **new and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy.** Just as we currently incentivize the creation of drugs, diagnostics and vaccines, **we should be incentivizing the creation, adoption, integration and effective use of novel datasets by local, state and federal public health agencies.** And as I have said so often in my responses to questions posed during this hearing, we can and must also ensure that personal privacy protections also exist -- **We can have both: personal privacy and access to the novel datasets needed** to stop the spread of COVID-19.*

36. Despite many structural challenges, Taiwan has fared better than many countries in dealing with the COVID-19 pandemic. Stanford Medical School documented 124 distinct interventions that Taiwan implemented with remarkable speed including community initiatives, hackathons, etc. Their “Face Mask Map” a collaboration initiated by an entrepreneur working with government helped prevent the panicked buying of facemasks, which hindered Taiwan’s response to SARS by showing where masks were available and providing information for trades and donations to those who most needed them, which helped prevent the rise of a black market.

What specific initiatives like this should we be implementing here?

**Inder Singh Response:**

*Kinsa has been in touch with senior representatives in Taiwan to roll out an early detection system in the country using our smart thermometers. Specifically, I recommend distributing 4.4 million Kinsa smart thermometers to households across the U.S. to effectively detect COVID-19 spread on a county-by-county basis. I believe this system is **essential infrastructure** today and will continue to be essential as our country faces worsening flu seasons and the possibilities of additional epidemics and pandemics. By knowing where and when outbreaks are occurring in real time, we can help our communities, the health care system and public health agencies direct resources effectively. Such an investment would most importantly save lives, while creating a significant return on investment, both in reopening the domestic and global economies and mitigating the impact of future outbreaks.*



## Sen. Baldwin

37. Emerging reports from many localities demonstrate that COVID-19 is having a disproportionate impact on African Americans and communities of color. For example, in my home state of Wisconsin, Milwaukee County reports that approximately 70% of those killed by coronavirus are African American, despite that community making up only 26% of the county's population.

We know this about Milwaukee County because the local government is proactive about collecting and reporting data on race and ethnicity. Reporting indicates that this disproportionate impact exists in places with significant African American communities, including Chicago, New Orleans, and Detroit. But a lack of consistent, quality data nationwide means we do not yet know just how sizable this disparity is, and what we can do about it.

While I am encouraged that we are drawing on the massive amount of data about Americans held by the private sector to support the COVID-19 response, I worry that it may not include and represent all communities equally. For example, if we use mobility data from mobile phones or particular apps to inform our understanding of adherence to social distancing requirements, I am concerned how it might affect the usefulness of the dataset if members of certain minority communities less likely to own such a device or utilize such an app.

For the members of our panel: how do you think "big data" can support efforts to strengthen our public health knowledge around COVID-19 and race, and how can we ensure that the methods and models through which "big data" supports our understanding of the epidemic take into account differences among communities?

## ***Inder Singh Response:***

*Thank you, Senator Baldwin. We agree that vulnerable populations, particularly those being disproportionately affected, need to be included in any data sets that analyze the spread of disease. However, disease spread typically occurs community-wide, and does not adhere to race or geographic bounds, and community transmission affects adjacent communities. As evidenced by COVID-19, this can happen within days. Thus, **we first and foremost need to at least understand how fast a disease is spreading in a geography**, and then we must always endeavor to understand how fast it is spreading within subcommunities. An optimized early warning system would do both. **We must start somewhere** if we are going to achieve the optimal state.*

38. I am also concerned about the impact of "big data" informing our COVID-19 response on rural communities. Again, I worry that some of these data sources may not be well-utilized in rural America – where connectivity is still a significant challenge – and thus may not reflect the reality of the pandemic in those communities. But I recognize that this information is vital to developing better predictive models that can inform our current response to COVID-19 and help us prepare for the future.

For the members of our panel: how does “big data” ensure that the different experiences of rural, suburban and urban communities are taken into account when informing models that may guide the COVID-19 response?

***Inder Singh Response:***

*Kinsa has undertaken initiatives like our school health program, FLUency, specifically to put our product in the hands of diverse communities, including rural, suburban and those living in inner cities. By securing corporate sponsorships to cover the costs of this program, **we have given away 200,000 Kinsa smart thermometers to primarily Title 1 schools** in under-resourced communities. I agree that an optimized early warning system would represent all population and demographics.*

*Kinsa’s intention has always been to create an early warning system for spreading illness. To accomplish this, we target three audiences:*

- 1. Larger households, where illnesses such as COVID-19 are likely to linger longer, leading to unwitting spread – in the case of COVID-19, once social distancing regulations are relaxed*
- 2. families in underserved communities, which typically seek care and treatment later, potentially spreading the virus before testing and treatment occur. Free thermometers donated to families in Title 1 schools through our school health program, FLUency, make up 20% of our existing thermometers in use today. We are expanding this program nearly 3X, to reach 4,000 elementary schools this coming Fall.*
- 3. Families of first responders who are most likely to come in contact with spreading contagious illnesses such as COVID-19*

*Furthermore, for every thermometer purchased through our website, Kinsa is donating 1 QuickCare thermometer to a family in need. We are also endeavoring to work with state and local governments so that they may purchase and distribute Kinsa thermometers to communities in need within their states.*

39. It is important that public health, and local public health departments in particular, have the data they need to map and anticipate hotspots for infectious disease outbreaks such as COVID-19 or overdose patterns in a community, including data that may be generated by the private sector. It is also important that local health departments have the capability to leverage this information together with that available through traditional public health surveillance efforts. For the members of our panel: how can the private sector coordinate data efforts with public health and ensure that local health departments have the necessary capabilities to make full use of these efforts?

### **Inder Singh Response:**

One gap in ensuring that local public health departments have the necessary capabilities to make full use of data, and evaluate innovative technologies and approaches to improving public health, is **better funding** dedicated to these areas. With more funding, local health departments can hire the right experts and carry out the work to better integrate data and evaluate innovations.

As it stands today, only a handful of federal agencies, and an even smaller number of teams within those federal agencies, are staffed, equipped, and funded enough to engage with new kinds of data or innovations. This risks that **many promising data sets and technologies will be overlooked.**

This has been our experience:

I started Kinsa with a mission to curb the spread of infectious illness through earlier detection and earlier response, primarily because I didn't see others -- whether in government or in the private sector -- leveraging innovation to solve this problem. In particular, we sought to aggregate **new kinds of data**: data on where and when symptoms were starting, how fast they were spreading, and how long lasting or severe they got. This has enabled Kinsa to predict flu spread well before CDC, including predicting the severity of the 2017-2018 flu season (as [cited by the NY Times](#) in early 2018), which killed 80,000 Americans and resulted in tents outside of emergency rooms.

While we have continued to succeed in our predictions -- as exemplified by [healthweather.us](#), a 2-week early warning system to first death from COVID-19 in 88% of states -- we continue to be overlooked by federal agencies, even after presenting peer-reviewed scientific literature supporting our work. I believe a key reason for this is that there is only a very small group of people at the national level making critical decisions for the country. This centralizes thinking, limits impact and reduces innovation. For the sake of the American people, we can do better.

**New and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy.** Better funding of local public health departments would decentralize decision-making and increase the number of teams evaluating innovative public health innovations and novel data sets. It would put power into local public health entities - entities that best know their communities and their specific needs. This would result in faster innovation adoption and better public health for our country.

40. In speaking with experts in Wisconsin working on developing and refining predictive models around COVID-19, I heard that while there is a significant number of both public sector and private sector data sources to inform models, the data is not consistently easy to obtain and incorporate. As we rely on real-time models to inform the COVID-19 effort, as well as look to prepare for future infectious disease outbreaks, it is important that data-sharing be as seamless as possible. For the members of our panel: what are ways we can strengthen the data-sharing infrastructure for government, public health, academic and private sector sources?

**Inder Singh Response:**

*Senator, I agree the framework needs to be improved. Indulge me for a moment while I build to a point:*

*First: A good friend and brilliant innovator, Andy Palmer, CEO of Tamr, often says that in today's world, data is a currency. If we look at data that way for a moment, there are a few things to be realized -- but first and foremost, **data is valuable.***

*Second: Analysis is only as good as the data inputs. If the data inputs are bad, the resulting analysis and data output is bad. As goes the old saying famous amongst MIT graduates and anyone who works in data analysis: "garbage in equals garbage out."*

*Given this, we must realize that **new and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy.***

*So that leads me to a conclusion: Just as we currently incentivize the creation of drugs, diagnostics and vaccines, **we should be incentivizing the creation, adoption, integration and effective use of novel datasets by local, state and federal public health agencies.** And as I have said so often in my responses to questions posed during this hearing, we can and must also ensure that personal privacy protections also exist -- **We can have both: personal privacy and access to the novel datasets needed** to stop the spread of COVID-19. This is critical to the health and welfare of our fellow citizens and country at large. It is during crises like this that we realize health and public health is a most important essential element of our lives and economy.*

*One way we can strengthen the data sharing infrastructure is by ensuring that local public health departments -- and not just well funded federal agencies -- have the necessary capabilities to make full use of data, and evaluate innovative technologies and approaches to improving public health. As it stands today, only a handful of federal agencies, and an even smaller number of teams within those federal agencies, staffed, equipped, and well-funded enough to engage with new kinds of data or innovations. This risks that **many promising data sets and technologies will be overlooked.** We must create departments in federal and local*

*government that can hire people with the necessary expertise and carry out the work to better integrate data and evaluate innovations. This will in turn lead to more rapid progress in the use of 21st century technologies for improving public health.*

## Sen. Tester

41. I'm very impressed by your program that distributes free thermometers to underserved communities. But I'd still bet that there are more of your thermometers in cities and suburbs than in rural areas and Indian country – the very folks that could be at greatest risk because of demographics and preexisting conditions. How can we protect our most vulnerable communities as we look to tools like yours to chart a public health response?

### *Inder Singh Response:*

*Thank you, Senator Tester. We agree that vulnerable populations, particularly those being disproportionately affected, need to be included in any data sets that analyze the spread of disease. And it is true that Kinsa's thermometer distribution tends to mirror population density, with more users in higher populated urban areas than in less populated rural areas. Due to the transmission dynamics of a contagious illness, **the more densely populated an area is, the more likely it is for illness to spread.** This is, for example, a reason why COVID-19 has spread so easily in places like New York, San Francisco and Seattle. We understand that rural areas and particularly Indian Country are not immune to the effects of COVID-19 and other pandemics. We would welcome the opportunity to work with tribal governments and the Indian Health Service to ensure our technology makes it into the hands of underserved populations across the U.S., particularly as the Kinsa app can serve as a triage system and telehealth conduit for populations with less access to medical care.*

*Kinsa continues to work towards ensuring our illness signal is a strong representation of the overall health of the U.S. population. We are collaborating with public health departments at the city and state level to distribute more thermometers to families and front line workers in areas hardest hit by COVID-19. By providing more thermometers to these communities, we strive to give the individuals in power the information they need to save lives. **We must start somewhere.***

## Sen. Sinema

42. Some states, including Arizona have limited testing capabilities and therefore limited testing. It is also widely reported that tests around the world have produced inaccurate results. How can we mitigate against inaccurate assumptions related to disease trends in situations in which we have limited or inaccurate data?

43. Many point to travel as a key factor in the spread of COVID-19. Contact tracing for travelers, specifically by plane, is a mechanism that can slow the spread of the virus. The data collected (full name, address while in U.S., email address, and two phone numbers) enables the government to contact individuals who may have come into contact with an individual who has tested positive. Once contact is established, individuals can start self-quarantining. What is the best way to balance the need for this information to slow the spread of the virus and privacy rights?

### ***Inder Singh Response:***

*Thank you, Senator Sinema. There are ways to use location based tracking and preserve personal privacy. I recently read a [brilliant comic in Kottke.org](https://www.kottke.org) that illustrated a contact tracing solution that would both protect personal privacy while also helping stop the spread of COVID-19. To explain briefly: contact tracing apps can work via Bluetooth and send a series of random "messages" of letters and numbers to any phones nearby. Because the messages are random and do not use GPS, they contain no personally-identifiable information. Any nearby phones with compatible contact tracing apps pick up and store the random messages sent by others in their vicinity. If any user tests positive for COVID-19, they or their doctor can input that information into the app. Any phones that hold a relevant "message" from the time the infected person would have been near them are notified that they were exposed. Throughout this process, no personally identifiable information has been captured, stored or shared.*

44. How can big data help resolve challenges within the manufacturing supply chain to spur increased production and distribution of needed testing, personal protective equipment, and other resources to address this pandemic?

### ***Inder Singh Response:***

***An early warning system is necessary infrastructure for our country. How do you stop an outbreak before it becomes an epidemic if you do not have knowledge of community spread in real time. The answer is you do not. You cannot. This has been abundantly clear with our challenges in procuring adequate PPE supply. Had an early warning system with enough lead time been in place, we could have known where to send our limited stock of PPE and of test***

*kits. Such information on where and when illness is starting, and where community spread is occurring, is vital in appropriately allocating limited supplies and manpower to the areas most in need of early intervention. It is during crises like this that we realize health and public health is a most important essential element of our lives and economy. We must not allow another epidemic to take American lives if we have the ability to stop it before it spreads.*

*We must realize that **new and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy.** Just as we currently incentivize the creation of drugs, diagnostics and vaccines, **we should be incentivizing the creation, adoption, integration and effective use of novel datasets by local, state and federal public health agencies.** And as I have said so often in my responses to questions posed during this hearing, we can and must also ensure that personal privacy protections also exist -- **We can have both: personal privacy and access to the novel datasets needed** to stop the spread of COVID-19.*

45. This pandemic has caused serious economic harm. Businesses of all sizes and their employees suffer as sales drastically fall or disappear altogether. State, tribal and local governments are under enormous strain as response costs increase and revenues drop.

How can big data assist in the better creation and execution of economic assistance programs like the Paycheck Protection Program, Treasury's lending facilities, business interruption or pandemic risk insurance, and state, tribal and local stabilization funds?



## Sen. Rosen

46. Germany's national disease control center recently asked their citizens to donate data collected by their fitness tracker. This voluntary initiative has consumers download an app on their phones and contribute health information such as pulse rates and temperature that is collected by fitness tracking devices anonymously. Using machine learning, epidemiologists can analyze this data to better understand the spread of the coronavirus across the country and detect previously unknown clusters.

What are the advantages and pitfalls in using voluntarily donated data to improve responses during a pandemic?

### *Inder Singh Response:*

*Thank you, Senator Rosen. In our experience, crowdsourced data, including through our own app, has several pitfalls, particularly around whether it is reliable or skewed. For example, people may voluntarily report when they have a fever, but are likely to not report anything if they are feeling unwell but do not have a temperature. This skews data significantly. The reason we created a connected thermometer & triage system was to ensure we get the full picture, including times when a user has a fever and when they may have other symptoms without a fever.*

*We believe that the data we receive, in the context of a medical guidance system & triage app, where the end user gets value for every individual input they provide and is context-specific, has much **higher integrity of data** than if you just ask people what symptoms they have. The science needs to be validated on any crowdsourcing app to ensure its usefulness. In our experience, context specific, triage apps that include connected medical products have the basic science behind them to do outbreak detection.*

How can we use donated data to support our response to this pandemic and future similar public health issues?

### *Inder Singh Response:*

*We can and should use donated data to do this. We just need to ensure that the science supports the use and value of donated data. For example, by running sound analysis between data donated on heart rate or temperature through a crowdsourced means and COVID-19 spread.*

What privacy guardrails are needed to ensure that this data is collected and analyzed safely and anonymously?

**Inder Singh Response:**

*Data collection, when done right, can create enormous benefits for individuals and communities, while protecting individual privacy. We too often focus on the collection of personal data, rather than on building protocols for the protection of personal data; we focus on the private benefits that accrue to companies that monetize data but often do not consider the possibilities of data use associated with public benefit; we focus on the right to privacy, but in doing so, we ignore another important right – the right to information. To be clear, I advocate for personal privacy. I also advocate for saving lives. And it is possible to do both. At Kinsa, we do not share individual data, we share real-time population health insights. Access to timely information is critical when it comes to public health, and the solutions essential for our country' welfare and health need to be advanced.*

What are the gaps we need to consider when analyzing such data?

47. Location tracking services serve as a powerful tool in understanding the movement of the coronavirus. Anonymized, aggregated data from GPS, Wi-Fi, and Bluetooth technology on our mobile devices can provide insights into how social distancing and shelter-in-place measures are changing people's behavior. A number of companies have come forward to help in the fight against the coronavirus, working to analyze and share these insights with governments on the local, state, and country level. They have stressed that the data collected is stripped of personally identifiable information.

But according to recent news investigations, researchers have developed a machine learning model that can correctly re-identify 99.98% of individuals in anonymized data sets with just 15 demographic attributes. In other studies, researchers used credit card meta data and with four random pieces of information were able to re-identify 90% of the customers.

To Mr. Graham Dufault and Mr. Inder Singh, what data security steps are your member companies/your company taking to ensure anonymized and aggregated data remain anonymized?

**Inder Singh Response:**

*Data collection, when done right, **can create enormous benefits for individuals and communities, while protecting individual privacy.** We have accomplished that at Kinsa. We do not share personal data, even if it is de-identified -- or anonymized -- data. With enough effort, even anonymized data has the potential to be re-identified. **We share population health insights** - the percent of people in a county who are ill. There is no way to identify an individual from this illness signal. While there are legitimate reasons for sharing de-identified data - for example, for research purposes or to evaluate the effectiveness of a health intervention -*

restrictions on attempts to re-identify the data should be incorporated into all agreements.<sup>12</sup> We need not sacrifice personal privacy for the sake of public benefit.

*We must realize that **new and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy.** Just as we currently incentivize the creation of drugs, diagnostics and vaccines, **we should be incentivizing the creation, adoption, integration and effective use of novel datasets by local, state and federal public health agencies.** And as I have said so often in my responses to questions posed during this hearing, we can and must also ensure that personal privacy protections also exist -- **We can have both: personal privacy and access to the novel datasets needed** to stop the spread of COVID-19. The tradeoff between personal information protection and providing information for society's benefits is a false one. **We can have both.** And we need both if we are going to successfully combat the second and third waves of COVID-19 coming which threaten the fabric of our healthcare system.*

48. The National Science Foundation (NSF) is the only federal agency whose mission includes supporting all fields of fundamental science and engineering. The research and educational programs backed by NSF are integral to the continued success of our country's innovation, supporting scientific discoveries that have led to new industries, products, and services. Since 2012, NSF has funded research on the emerging field of data science through its BIG DATA program. Now, NSF's larger program – "Harnessing the Data Revolution" – will support research, educational pathways, and advanced cyberinfrastructure in the field of data science.

Given NSF's leadership in data science research and development, what role do you think NSF can play in leading public-private partnerships for increased research on big data that could help address the COVID-19 crisis or future pandemics?

### ***Inder Singh Response:***

*We would welcome working with the NSF. We believe it is critical that there be a government entity that is embracing new innovative technologies and private sector initiatives that are getting to the heart of how we do outbreak detection, prediction and monitoring. As it stands today, only a handful of federal agencies, and an even smaller number of teams within those federal agencies, are staffed, equipped, and funded enough to engage with new kinds of data or innovations. This risks that **many promising data sets and technologies will be overlooked.***

*This has been our experience:*

<sup>12</sup> *We as a company have only shared de-identified data in 4 instances, and only for the purposes of research or program evaluation. This research has resulted in a number of scientific, peer-reviewed publications. In these instances, we are careful to always add clauses to ensure there is not an effort to re-identify the data, and use of the de-identified data is limited and narrowly defined.*

*I started Kinsa with a mission to curb the spread of infectious illness through earlier detection and earlier response, primarily because I didn't see others -- whether in government or in the private sector -- leveraging innovation to solve this problem. In particular, we sought to aggregate **new kinds of data**: data on where and when symptoms were starting, how fast they were spreading, and how long lasting or severe they got. This has enabled Kinsa to predict flu spread well before CDC, including predicting the severity of the 2017-2018 flu season (as [cited by the NY Times](#) in early 2018), which killed 80,000 Americans and resulted in tents outside of emergency rooms.*

*While we have continued to succeed in our predictions -- as exemplified by [healthweather.us](#), a 2-week early warning system to first death from COVID-19 in 88% of states -- we continue to be overlooked by federal agencies, even after presenting peer-reviewed scientific literature supporting our work. I believe a key reason for this is that there is only a very small group of people at the national level making critical decisions for the country. This centralizes thinking, limits impact and reduces innovation. For the sake of the American people, we can do better. There are technology solutions coming out of the private sector -- such as Kinsa's, or those from NSF I-Corps Grant recipient Elektra Labs -- that are promoting safe monitoring and need to be accelerated to help stop the spread of COVID-19 and other outbreaks.*

*I also believe we must ensure that local public health departments have the necessary capabilities to make full use of data, and evaluate innovative technologies and approaches to improving public health. With more funding, local health departments can hire the right experts and carry out the work to better integrate data and evaluate innovations.*

***New and novel data sets that help us do better analysis -- such as predicting where and when outbreaks like COVID-19 will spread -- are of immense value to our public health system and our economy.*** Better funding of local public health departments would decentralize decision-making and increase the number of teams evaluating innovative public health innovations and novel data sets. It would put power into local public health entities - entities that best know their communities and their specific needs. This would result in faster innovation and better public health for our country.