Contagiousness and Spread of SARS-CoV-2

[Announcer] This program is presented by the Centers for Disease Control and Prevention.

[Sarah Gregory] Hi, I'm Sarah Gregory, and today I'm talking from home with Dr. Ruian Ke, a scientist at Los Alamos National Laboratory in New Mexico. We'll be discussing how quickly COVID-19 spreads and possible steps to control it.

Welcome Dr. Ke!

[Ruian Ke] Thank you for having me here, a pleasure to be here.

[Sarah Gregory] What's the difference between SARS-CoV-2 and COVID-19?

[Ruian Ke] SARS coronavirus 2 is the virus that causes coronavirus disease, which is short by COVID-19. I used to…I have been working in HIV, and this is very similar to the comparison between HIV and AIDS; HIV is the virus that causes AIDS. So, one is the virus…the other one, it refers to the disease it causes.

[Sarah Gregory] Ok, so COVID-19 is, is the disease.

[Ruian Ke] Exactly.

[Sarah Gregory] What initially happened in Wuhan that made people aware of something unusual going on?

[Ruian Ke] So, there were 41 cases of pneumonia of unknown etiology that alerted people. This is what's reported to World Health Organization, WHO, by the Wuhan Municipal Health Committee at the end of December 2019. And a lot of cases were linked to the Huanan seafood market. And this later on confirmed…this pneumonia was later confirmed to…to be caused by the virus. In late January, it was made aware that this virus can transmit from human to human. That really alerted the world the danger of this virus.

[Sarah Gregory] What's mathematical modeling and how is it used to figure out how a virus will spread?

[Ruian Ke] The types of modeling we use is to use differential equations to describe how infectious disease would spread among a population over time. By fitting the model to existing data, we would be able to figure out the parameters in the model where the model has a mechanistic description of the spread of the virus. In that case, we can use what we know already to project what would happen in the future and how an intervention would impact on future epidemic trajectories.

[Sarah Gregory] Can you give…maybe this is too complicated…but could you give an example of what a mathematical model would look like?

[Ruian Ke] Sure. The model we use is so-called as SEIR, susceptible, exposed, infected, and recovered. So, basically describes how an infected individual would go from susceptible to the virus, to become exposed to the virus, then become exposed to infected, and from infected to recovered individuals. Basically, we write down equations to describe how a population of individuals goes through these different stages, and that would help us to portray how the virus would spread in a population.

[Sarah Gregory] Okay. What makes studying a newly emerging and fast spreading disease, such as COVID-19, challenging?

[Ruian Ke] Sure, there's a number of challenges for…for researchers from a modeling perspective. The most challenging part is the limited amount of data during early outbreaks and the answers to data corrections. This is because of many unknowns associated with the new emerging pathogen. The questions are usually raised about what type of data we want to use and how to appropriately integrate those data based on existing knowledge. This becomes very challenging to model an emerging pathogen like COVID-19.

[Sarah Gregory] Your study was about how contagious COVID-19 is and how rapidly it was spreading. How did you go about finding these things out?

[Ruian Ke] Yeah, that's a really great question. I, as I mentioned, due to these uncertainties of data collection and also a lot of complications in data collection in Wuhan, the epicenter, we took an indirect approach. The idea is that it's often better we can infer what was happening in Wuhan, the epicenter, and the Hubei province, where Wuhan is, from data collected outside of Hubei. The reason why we try to use data outside of Hubei is because, in Wuhan, things happened so quickly in the beginning, we didn't know what the causes of the disease was. It took a while to, really, to get to know this is caused by a virus, and it took a while to give out diagnostic kits to…to detect and confirm an individual infected by the virus. And over time, quickly, the virus already infect a lot of people in a short period of time and healthcare systems got overwhelmed. As the data collected during that period when healthcare system is overwhelmed is…is hard to integrate, as well. So, this is why we took an indirect approach to use data outside of Hubei to infer what was going on inside Hubei.

The, the major advantage using data collected outside of Hubei, one, is that when the amount of infected individuals showing up in the other provinces, the…we already know how to diagnose those patients and also the number of infected individuals are, in general, are generally low, the healthcare system isn't that overwhelmed. And those are the basis of a lot of the uncertainties I mentioned for data collected in Wuhan. So, the figures have come using data outside of Hubei and trying to infer what was in Wuhan, minimized a lot of the biases in both data collection.

But I want to also mention that, in addition to the efforts to try to minimize biases in data, we altered two different modeling approaches, which constructed two different kinds of models to use different aspects of the data we collected to make inferences. And, interestingly, both models reached similar conclusions that the virus spreads extremely fast in Wuhan. So, that adds a lot of confidence of the conclusion of our…that is, again, back to your question about the challenges in understanding a rapid-growing epidemic, we really tried to use multiple approaches to see better if we can get similar conclusions.

[Sarah Gregory] What were the various kinds of data that you used?

[Ruian Ke] Yeah, we practiced a wide variety of data sets. One key data set we want to…we collected and really want to emphasize or mention that make our study reliable and interesting is that we collected a very high level of mobility data from a online server called Baidu Qianxi. Basically, this server summarizes daily train travel between cities and provinces based on mobile phone positioning data. And, I want to mention this server is the most likely used server in China in terms of positioning services. And this really provides a high-resolution understanding of travel**,** where we can compare this during the period when the COVID-19 spread across all over China. And we believe this data really enable us to infer the…the epidemic growth in Wuhan from data outside of the Wuhan and Hubei province.

Another data set, which is very unique, that we crafted, is the travel information of the first or the first few persons who were confirmed of COVID-19 infection outside of Hubei province. Those information gave us an understanding or gave us a estimate of how quickly the virus spread across China, as I will mention, at the rate of how quickly cases show up in different provinces in China gave us a hint of how quickly epidemic growth in Wuhan reproduces, is another key data set we collected.

[Sarah Gregory] Early on, you apparently gathered documents and news sources from CDC China and Chinese official websites. What did you do with this information, and how did you turn it into a data set?

[Ruian Ke] There were enormous efforts on data collection during January we got really interested in and thought this could be a problem. So, we gathered all this information from China CDC to local health commissions, where they make a lot of important epidemiologic information publicly available. One thing I want to also mention that the benefit is from the multicultural background of our team. Those information were first published in Chinese. We did a lot of data collection, collection of information and translation of those information to English to be analyzed. We almost had firsthand information by directly translating those documents into English and we compiled those information and categorized each case report according to their dates of travel to Wuhan, dates of symptom onset, hospitalization, and case confirmation. All of these are equal to epidemiological information that were later on used for our estimation.

[Sarah Gregory] After you pulled all of this early data and information together, what did you find?

[Ruian Ke] We found this extensive data set were able to estimate several key epidemiological parameters, including the duration of the incubation period, which is estimated, according to our data set, was, has a mean of 4.2. It's roughly consistent with other studies showing the median of the incubation period of four days and a mean incubation period between five to six days. And, importantly, I want to mention that this is during mid- and late January. Those, the estimates from other studies weren't available. We were able to use the estimation through health data in public available data sets to inform the mathematical model we abstracted to estimate those rates. So, we were really excited about those findings, to have those key parameters and values estimated.

[Sarah Gregory] So, are you saying the estimated five to six-day incubation period is still holding true?

[Ruian Ke] I think so. The estimate for the…of those studies all have confidence intervals, I think, four to six days are all the same confidence intervals estimated across different studies. Or…if you…I, I think those are, in general, consistent with each other and hold true as for now.

[Sarah Gregory] Well, that's good to know, that's better than 14 days, I guess.

[Ruian Ke] So, I want to add that, since you brought up, those estimates are the mean of the estimates of, there's some study reported in, is technically not the…I think that 14 days incubation period was the upper bound, or most of the people would accept missing 14 days—it's definitely not the incubation period. What we estimated here was a mean estimation.

[Sarah Gregory] Were there potential problems with the early data?

[Ruian Ke] Yeah, one caveat, as pointed by one of our reviewers, is that the individuals who were reported at that time are often to match individuals and many of them may have severe symptoms, and that is why they end up being in hospital. Others are…through transportation hubs, so there's some potential bias in terms of these estimations, which, too, was characteristic in more severe…to individuals having more severe symptoms.

[Sarah Gregory] How did you estimate the growth rate of the outbreak in China in January? And tell us what an "R naught," is. We've been hearing a lot about that in the media lately.

[Ruian Ke] Yeah, sure. Definitely. Let me answer the question about R naught. Strictly speaking, of the formal definition of R naught is a reproductive number, and is formally defined as the average number of individuals that the first infected individual infects when this first individual is introduced into a fully susceptible population. But I want to clarify a lot of confusion about R naught. This quantity is the average number of individuals when the first individual introduced in a fully susceptible population. So, if we calculate how many individuals, in fact, during the epidemic, strictly speaking that should be defined as infected, are infected, our infected reproductive number was not R naught. So, formally, it should be a number of when an individual introduced in a fully susceptible population as a measure for transmissibility. To estimate this number, usually we would first estimate the epidemic growth rate. During the very early outbreak, the…the assumptions are really, really bad. During early outbreak, the susceptible population is not depleted as the growth rate truth reflects the…the infection or the transmissibility of the virus when it is first introduced in the susceptible population.

So, about your second question, about how did we estimate the epidemic growth rate. What we did was really to—again, back to your great question about uncertainties—to estimate a…a rapidly growing novel pathogen, we tried to minimize potential biases in dataset and also model construction. We took two alternative approaches. The basic idea is…can be rooted in sciences or physics where, if the epidemic grows exponentially in Wuhan, then if we assume people travelling from Wuhan to other provinces are almost like a random draw, then how quickly will we detect influx of the cases across different provinces would reflect how quickly epidemic grows in Wuhan. I think by collecting data of how quickly the first case showing up in each provinces, and how quickly the cases showing up in these different provinces, can be inferred how quickly the epidemic grows in Wuhan.

[Sarah Gregory] And I believe there was some other evidence of growth rate, right?

[Ruian Ke] Yeah, absolutely. This is also, I think, is important in terms to cross-validate our findings. We not only look for…for…not only look for data that not only is the two alternative approaches to be sure the growth rate we estimated to are accurate, we also looked at other evidence that we did not use for our modeling trend to confirm the conclusion from our inference. So, what…one important set of data is the growth of the death counts over time in late January in China. We looked at how quickly death count grows, and we realized it grows at similar rates as we estimated for the growth of the epidemic. We think this is a strong indication of the…the reliability of our conclusion because the growth of the deaths is…should reflect the growth of the epidemic with some lag, is mainly because of the…the…the lag between infection and death. And…but how quickly the death grows should be roughly reflecting how quickly the infection grows early on. So, we've found that the growth of the death counts increases directly at a rate as we estimated for the growth of the epidemic. That is one important dataset that cross-validates our findings. The other important dataset was later published by China CDC. They

showed the epidemic curve based on symptom onset, and those curves again showed that the epidemic grows really, really fast in January 2020.

[Sarah Gregory] Ok, I realize that this is an extremely technical process, this mathematical modeling and everything you've done. So, would you tell us about the implications you found for intervention strategies?

[Ruian Ke] Yes, sure. Absolutely. Based on much higher estimated growth rate, we calculated where R naught has a median of 5.7. This is much higher than the previously estimated R naught, which is between two and three. The first implication I want to mention is the importance of the calculated value of R naught in terms of herd immunity, which we heard a lot recently in media. We can directly accomplish a herd immunity level required to stop transmission. This is calculated as one minus one over R naught. Just to give you an example, for an R naught between two and three, the herd immunity threshold is between 50 and 60 percent. So, we need 50 to 60 percent of the population to be immune to the virus, to stop virus transmission. But for our R naught of 5.7, this threshold increases to 82 percent, so there is a much larger fraction of the population needed to be either vaccinated or infected so that the herd immunity will achieve to stop the virus transmission.

The other implication for such a rapid growing virus and has…is having such a high R naught is that, we found if as little as 20 percent of transmission through and by the individuals who are not identified and rapidly isolated and their contacts are not quarantined, the transmission would be extremely hard to control. We found that…at that time, we didn't actually know how many people, was the fraction of people that are unidentified by this surveillance system. Right now we know there are a large number of people that are not detected or identified by the surveillance system. So that really means early strong interventions, such as social distancing, is actually pretty needed. Otherwise, it would be hard to control, the current technology, with respect to contact tracing and isolation. I think those are the two most important implications of our work.

[Sarah Gregory] Let's go back over this, and would you summarize for us what does your study tell us about the incubation period, the time from first symptoms to hospitalization, and how contagious the virus is?

[Ruian Ke] Yes, sure. What we found was the mean incubation period has a mean of 4.2 days, and the time from first symptom to hospitalization is…is time varying. In the beginning, the time from symptoms to hospitalization is relatively long, it's about five to six days, in the beginning. And then, in late January, this time becomes shorter. It only takes about one to two days to get hospitalized after they develop symptoms. And for the contagiousness of the virus, we found this virus is very contagious. It's much…it spreads much quicker than estimated initially in the literature.

[Sarah Gregory] How does this virus differ from the 2003 SARS virus?

[Ruian Ke] Great question. From our study, it definitely shows it is more contagious. It spreads much faster than the previous SARS viruses. I guess as the epidemic goes on, we'll know much more about this virus. Now, SARS-CoV-2 can transmit asymptomatically or before symptom onset. This is very different from 2003 SARS virus. And also, this virus causes a wide variety of disease severities. Most of the people infected by the virus have no symptoms, some develop mild to severe symptoms, and some develop very critical conditions or even die. However, for

2003 SARS viruses, as far as we know, most people developed very severe symptoms. And the 2003 virus is more transmissible after symptoms are developed.

[Sarah Gregory] I know this is probably just a personal opinion, but do you think it's possible to contain this virus now?

[Ruian Ke] This is a very hard question to answer. It really depends on the…the efforts we…or the control efforts that…that is implemented in a particular area. What we found was this virus is hard to contain. However, for example, from China and South Korea it looks like, when strong interventions are first implemented for certain periods of time, they can control the virus at the low level. From my own perspective, I think it can be contained with substantial control efforts for a substantial period of time.

[Sarah Gregory] How does this study advance our public health awareness, and how should it impact approaches to containment?

[Ruian Ke] I think what we show is the virus can spread extremely rapidly. We should really be prepared in advance. Otherwise, then, if we react slowly, we give the virus room for…to spread rapidly. The other implication that we want to be aware is that, given the rapid spread of this virus, when we think about relaxing a lot of the social distancing efforts, we want to be very cautious. Otherwise, the virus can spread very, very quickly and reverse the efforts we already put in and the sacrifices that we already put in to slow down the curve to…to reduce the infection. So, we have to be very cautious, given the high risk of spread of this virus.

[Sarah Gregory] Tell us about your job…where you work and how you're involved in working with this virus?

[Ruian Ke] Sure. I work in Los Alamos National Laboratory. The division we belong to is called the Theoretical Division, and the group I was working with is called the Theoretical Biology and Biophysics. A major effort in this group was to use mathematical models to understand how virus spreads and how virus infects individuals and what the implications of drug treatment, for example. Due to the rich culture in this group, I got really involved in early January when this virus was first identified and we paid a lot of attention to the development of this virus. And we realized there is a potential threat to the world in early January, so that's how we got involved in working with this virus.

[Sarah Gregory] On a completely personal note, are you at home and what are you doing to stay positive?

[Ruian Ke] Yeah, I'm working at home now to reduce contact with…with other people. I think one thing that keeps me positive every day is the urgent need to find solutions to the global crisis this virus causes. I think…thinking looking at the news and looking at people suffering… people's lives are lost. Those tragedies really have been a huge motivation for me to wake up every day and concentrate on work and try to contribute to the understanding and to the fight of this virus. And hopefully we get the solution to end this crisis soon.

[Sarah Gregory] Well, thank you so much for taking time out of doing exactly that, to talk with me today, Dr. Ke.

[Ruian Ke] Thank you for having me here!

[Sarah Gregory] And thanks for joining me out there. You can read the July 2020 article, High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2, online at cdc.gov/eid.

I'm Sarah Gregory for *Emerging Infectious Diseases*.

[Announcer] For the most accurate health information, visit cdc.gov or call 1-800-CDC-INFO.