Cecile Viboud

Behind The Mask

September 15, 2023

Barr: Good afternoon. Today is September 15, 2023. My name is Gabrielle Barr, and I'm the Archivist at the Office of NIH History and Stetten Museum. Today I have the pleasure of speaking with Dr. Cécile Viboud. Dr. Viboud is a Staff Scientist in the Division of International Epidemiology and Population Studies (DIEPS) at the Fogarty International Center. Today she will be speaking about the many studies she has been a part of for COVID-19. Thank you very much for being with me.

Viboud: It's a pleasure to be here.

Barr: Will you please share how you first learned of SARS-CoV-2 in December of 2019?

Viboud: I think it was through collaborators working on the viral sequencing and phylogenetic analysis. Through Twitter, I heard about this new pneumonia syndrome in China. Very early in January [2020], there were a few virtual meetings organized by WHO [World Health Organization] and CDC [Centers for Disease Control and Prevention] to start to worry about this thing and try to learn some of the characteristics of the pathogen from what could be gleaned from China.

Barr: That's so interesting—from Twitter.

Viboud: Yes, yes. In the virology world there are a few groups for Twitter, but there's also a group that's called Virological where they sort of monitor every little fret out there. Anything unusual is commented on by the entire virologist community. Those are good channels to be on when you're interested in these kinds of things.

Barr: Will you please discuss one of your early studies where you and others compiled and disseminated epidemiological information on COVID-19 from social media networks and the news, and what the value of this approach is especially in the beginning of a pandemic?

Viboud: Yeah. When I started working on this, again in very early January, there was very little individual level data—data on each patient—that was available. Those are data that are very precious when you want to build mathematical models, because those are data in which you can both understand the age distribution and if there's anything unusual, and if there is an age group that's more affected than another one. And that was the case, right? I mean, we quickly understood that older folks were more affected. If you also have chance of transmission, so clusters of transmission, you can sort of see where the epidemic goes from person to person and how long it takes. That's called a serial interval, which is a parameter that's very crucial for transmission models and just having a sense of the general epidemiology of this disease. Very little information was available, but at the same time, we knew that there was a lot that was published online, in particular from China, who focused on social media for the medical practitioner. By chance, we had a postdoc in our group who was Chinese. Because of course, all of this information is in Chinese, and he had just returned from China. It was obviously his home country, and he was very interested in this outbreak. He and a research assistant in our group started compiling all of the data that was available online. Every time there was a new case, a local doctor would say, “This is a 37-year-old woman, she comes from this country, she has these symptoms, etc.” You can

collate this information in a tabular format in a structured way and also translate it, of course. Everything was translated into English. Quickly, we arrived at several hundred patients. We made this database available online, and it was the first publicly available data of this kind. A lot of people then did the same thing, and all of these eventually got merged into a general dataset. But it was both to try to prove that you could use social media to use this data and then also to do something about the public good and put data out there.

Barr: Early in the pandemic, you were part of a team that used a global metapopulation disease transmission model to project the impact of travel limitations on the national and international spread of COVID. Were you surprised by the slim effect it had on the pandemic trajectory overall?

Viboud: Yes. It is a very obvious thing to do, and it's been done time and time again in the cases of past epidemics or pandemics, just to try to close borders. But actually, there is a lot of modeling history that had shown that this would not be effective—a lot of work around pandemic flu, because the same question arose in 2009, when we had the swine flu outbreak. The same type of global models with travel connectivity were used, and they said unless you can reduce 99% of travel globally, which is almost impossible, and you do it very, very quickly, it's not useful, because the epidemic has already disseminated by the time you try to close borders. And in this paper, it's shown that if you close borders, you delayed the epidemic by one to two weeks in terms of magnitude. But in the grand scheme of things, it's not doing very much. It was not a surprise, because we had seen something similar with similar viruses in the past, but it's always a message we need to get out every time.

Barr: Can you discuss your efforts to look at the case fatality risk of the first pandemic wave of COVID and China?

Viboud: Yeah, to be honest, I don't remember this paper very closely. I think we also looked at data reported by the provincial level CDC [Chinese Center for Disease Control and Prevention] as well, or maybe a little bit of social contact data. The main thing that you worry about when you estimate the case fatality risk is that you're not seeing all of the deaths right away, because you're seeing cases, but deaths will lag and tend to occur later. And so, a lot of the estimates early in the pandemic are underestimates of the case fatality rate. This is what this model tried to do—sort of impute what you're not seeing yet.

Barr: One of the important papers you were a part of was based on patient and contact tracing data from Hunan, China in 2020. What was revealed about transmission heterogeneities, kinetics, and controllability of SARS-CoV-2 at that point in time?

Viboud: The main message of the paper—and it was using very detailed and unusually high-resolution data from contact tracing in China, the type of data we never got in the U.S., unfortunately, and in most countries—was that you could trace the timing of transmission of COVID-19 from one patient to the next relative to the timing of symptom onset. What we showed there is that half of the transmission occurs before symptoms appear—so when the index case doesn't know that he or she is infected. Because of that, it's very difficult to control SARS-CoV-2 because as soon as you have some symptoms, you want to test yourself. Maybe you isolate. Before that you're actually still quite infectious, but there's no way for you to know that you're already infected. We also saw a lot of heterogeneity in the number of contacts between individuals, but I think that's seen in a lot of studies. You have people who are social butterflies and meet with a lot of other folks. And others would essentially just stay at home. It also varies across age groups—children tend to have a lot of contacts. The last thing that was quite interesting in this study is that we did it throughout the period of lockdown and isolation—very strong isolation—and you could see a very sharp decrease in contacts there. Essentially, the only opportunity to transmit is within the household before the household is isolated and all of the other contacts are sort of removed, so you clamp down on transmission very fast.

Barr: During this period, you were also part of a study that looked at infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under the period of intense contact tracing in Hunan, China. Will you just say a little bit more about what was learned from that angle?

Viboud: Yeah, so this was sort of a companion study to the one I just talked about. It was using the same contact tracing data but looking more specifically at whether some age groups were more or less susceptible. Early in the pandemic it was decided that maybe somehow children were refractory to COVID-19, perhaps because they didn't have the right receptors. We wanted to look at that more closely. And we didn't really see strong evidence of that. Now it's been debunked. Children can get infected. Obviously, they don't have severe disease at all, so it's a mild infection, but they can get infected. Then we also see differences in infectivity, depending on potential for transmission and depending on contacts, but not so much on age groups.

Barr: In the same vein as contact tracing and extreme isolation in China, will you share a little bit about your findings from devising a set of contact surveys performed in four areas in China—Wuhan, Shanghai, Shenzhen, and Changsha—during the pre-pandemic lockdown and post-lockdown periods, in regard to changes in contact patterns and how that affects transmission?

Viboud: Yes. This was again a very unique study in the sense that there had been contact surveys done before COVID in a couple of cities, and collaborators in China at Fudan University in Shanghai were able to do the same type of survey during the lockdown period—even though, obviously, any kind of research study was pretty difficult by that point. You could really compare, in the same city, pre-COVID 2019 contact patterns versus spring of 2020, which was the period of the big lockdown. You could see, as expected, that all of the contacts that are outside of the household disappeared. Typically, the contact matrices are what we call very assortative. One age group will be in contact a lot with the same age groups. That's particularly because children are in schools with the same age groups. For working adults, there's a little bit more spread, but we tend to interact with other working age adults, and older folks tend to cluster together. All of that disappeared, and the only thing you had were contacts within the household. Based on the survey, we could also infer what the expected transmission of COVID-19 would be if we resumed some of the contacts—for instance, if we resume school cycles, but just that and nothing else, would that be enough to sustain transmission or not? Having this contact survey was quite good.

Barr: Another interesting study that you did was a population-based study that looked at health seeking behaviors of individuals in Wuhan, China, who are experiencing acute respiratory infections. What were the outcomes of the surveys, and how does that influence disease burden, clinical severity, and resource allocation, as was as the subsequent intention of the study?

Viboud: I'm blanking a bit on this study.

Barr: There are so many, it’s hard to remember all of them. In looking at the role of social distancing and mitigating the spread of COVID-19 in South Korea from January 20, 2020, the first confirmed case in that country to March 16, 2020, you mentioned some of the sources that were used to compile this information and some of the finding.

Viboud: Yes. We used data to measure mobility. We used data from the commuting system—from the Metro—where they record who takes the Metro or not. That gives you an idea of a volume over time so that was quite useful. A lot of people have used GPS-based mobility data for the pandemic, but this is very aggregated. The

Metro data is very walk driven, so it's a pretty good indicator. Then we use publicly available case count as reported by the Korean Ministry of Health so you could follow the decline in transmission as mobility in the Metro declined. Some provinces outside of Seoul were a little bit more disconnected with the rest of the country and sort of on their on their own epidemic dynamics. It was quite interesting.

Barr: What is it like with countries having that type of data available? I'm sure some countries have more out there than others. In a country like the U.S., how much is private, or kind of all over the place? What's it like to track down that kind of information?

Viboud: Yeah, it's been very tough. This is not the first time we encountered this. In the 2009 pandemic, we also had the same issue. We know our healthcare system is very disparate, so having a good idea of who is tested and hospitalized is very tricky. To try to respond to that, the CDC put in place a lot of different surveillance systems for testing. They also created a very nice dataset that's still going on right now and that we didn't have before, which is this HHS [Department of Health and Human Services] Protect Public Data Hub, which tracks every hospitalization due to COVID, and now due to flu as well, across the U.S. Every hospital has to report that even though they might be a private hospital, and that's been really good. Our death system is actually very good in the U.S. with vital statistics. The one issue is that it's always been lagging in time a lot because we're a federal system, so it used to take three years to get any data, which is not admissible in a pandemic. There's also been really good changes for vital statistics. They've really made an effort to release publicly available data every week throughout the pandemic, and some of that is still ongoing now. A lot of modernization. I think we need a lot more, but it's difficult with a federal system. It's difficult because we don't have universal health care, so you have to think of different systems.

Barr: Early in the pandemic, it was difficult to gauge COVID-19 as a cause of death in the United States, due to difficulty with testing and lack of reporting, as you just said. Will you discuss your part in an observational study that looked at excess deaths from COVID-19 from March of 2020 to May of 2020—so very, very early in the pandemic—and explain why there is substantial variability between states and the difference between official COVID-19 death and estimated burden of excess deaths?

Viboud: Yeah, so as you said, this was very early on in the pandemic, during what we call the first wave. The first wave was quite sporadic across the U.S. A few states were hit quite early—New York was sort of the poster child for that, probably because their connectivity with other places, including Europe and China, is very high. Most states in the U.S. didn't really see very much during that first few months of 2020. Also, there were a lot of differences in testing and in coding of deaths. The excess mortality approach is an approach that doesn't rely on coding or testing. It's just fitting a statistical model to past deaths in prior years before COVID-19, using this as a baseline, and then everything that's above the baseline is excess. If you know that something very big had happened—that could be a pandemic or a heatwave or a hurricane—then you can probably safely attribute the excess to that event. When we did this and then compared it with the deaths that were coded as COVID-19 through testing, we saw quite a bit of differences. Most states didn't have enough tests, and as you said, it varied a lot just because of propensity to test. That actually didn't change very much during the pandemic. There was more testing across the board in general, but still heterogeneity in testing.

Barr: Will you discuss how you used a time series approach and generalized additive models in a later similar study that evaluated the direct and indirect mortality impacts of the COVID-19 pandemic in the United States from March 2020 to January 1, 2022?

Viboud: Yeah, so this was kind of a follow up study taking a slightly different approach. This time, we're looking at four waves of the pandemic, essentially just before Omicron. We had two waves of the original strain, the Alpha wave, and the Delta wave. The pandemic probably caused a lot of death through direct SARS-CoV-2 infection, but there was a question as to whether there were maybe other indirect types of death that were not directly related to COVID. At the same time, there was all of this worry about the healthcare system being completely saturated, so maybe we couldn't take care of cardiovascular infection. There were also reports of opioid deaths really rising. This is what we can think of as the indirect impact of the pandemic. For this approach, as I said before, we sort of did a baseline for past years. We let it run for the pandemic, and then we looked at weekly excesses. With the regression approach, we tried to attribute those weekly excesses to either SARS-CoV-2 circulation or the impact of interventions, because we had indicated how strong interventions were every week in the U.S., and then also healthcare saturation. We saw that across the U.S. 84% of excess deaths were COVID-19 related as direct impact. That's the major thing that we see that's important, but also, the 16% of indirect effects, at the scale of the U.S. where we had about a million deaths by that point of the study, is quite a lot. This is even more pronounced in the younger age groups that are not traditionally affected by COVID-19. Then looking at causes of death, we saw that in particular, accidents, opioids, and homicides had risen a lot—up to 20% during COVID-19, and in periods where COVID was not really circulating so you cannot attribute it to COVID. But it's just mental health and the traditional ways to deal with these things were crumbling or not quite there, because people were doing other things. Those indirect effects are also important to consider trying to prevent for next time if we have these big lockdowns again.

Barr: Will you speak about what types of models and analyses were used to evaluate trends in risk factors and symptoms associated with SARS-CoV-2 and rhinovirus test positivity in King County, Washington from June 2020 to July 2022? What were some of the relevant takeaways from this study?

Viboud: It's a big study. This was a very large study. King County is the greater Seattle area in Washington. A few years before the pandemic, there was a lot of effort put into surveillance in Seattle. Actually, the idea was to try to beef up surveillance and see if we could prevent a pandemic. And then the pandemic happened kind of randomly, but the study was in place to do a lot of testing. It was testing in the community with a nice system where people, when they have symptoms, go online and they request a test kit. The kit is shipped to them by rapid courier; they get it within 24 hours. They swab themselves, put it back in an envelope, and the courier gets it back to the lab. Then it's a PCR [polymerase train rection] test. The nice thing is that the study did a multiplex pathogen where they tested for 26 different pathogens, including COVID, flu, rhinovirus, and a lot of other things. In addition, there's a lot of sequencing for that study. That's for the community part. The study collaborators also worked closely with hospitals in the Seattle area that provide residual samples or respiratory samples from people who were seen at this hospital and again, go for these 26-pathogen testing and sequencing. You get a very good idea of what is circulating both in the community—people who might not necessarily seek care—but also in the hospital. That's the most severe part of the disease pyramid. This study was conducted before, but also during, the COVID-19 pandemic. Actually, during six months, it was kind of the only way to get tested in Seattle. We also have a lot of patient metadata, so you could look at risk factors. It was interesting to compare with rhinovirus, which is the only other pathogen that kept circulating throughout the pandemic. Most of the other pathogens disappeared for a while because contacts were so low, but rhinovirus stayed so we could see, for instance, that being very young or under five years old was a risk factor for rhino, but not for COVID. We see that travel was a risk factor for COVID throughout the pandemic. We could also see differences in the risk of infection with the Delta variant versus the Wuhan strain. You could also look at how people were vaccinated and whether they were positive or not. You can get estimates of vaccine effectiveness. We could see how that declined for the Delta wave, and how it improved when you were boosted. You can really look at changes in the epidemiology of the pandemic throughout time for this type of data.

Barr: In the systematic review and metanalysis of serological evidence of human infection with SARS-CoV-2, will you discuss your findings that antibody-mediated herd immunity was far from being reached in most settings during the latter half of 2020 in spite of mitigation efforts and how estimates of the ratio of serologically detected infections per virologically confirmed cases across WHO regions could help provide insights into the true proportion of the population infected from routine confirmation data? How was this implemented with later models?

Viboud: We were particularly interested in serology in Africa, because there's been a lot of questions about the disease burden of COVID there—there's very little data on hospitalization. In that case, it was serology—at least we know who has been infected. It doesn't tell you who got clinically ill with it or died or anything, but it at least gives you an idea of the pervasiveness of the disease. We saw that serological attack rates were very high in Africa, compared to, for instance, the U.S. or Europe, even in the first year of the pandemic. It really meant the disease had been there. When you compare the infections with reported deaths, you had very different ratios across countries because of very large differences in testing rate. Just to speak to the role of serology in disease models, it's been used quite a lot to try to calibrate disease models, because that's, again, a really good measure for how many infections and how many people have seen the virus. If you just look at reported cases or hospitalization, you know that it's always under reporting, right? Or even people who would never get symptomatically infected. That can really bias the trajectory of the epidemic and the projected dynamics. Serology is quite important right now.

Barr: What are the limitations? In another paper you wrote that "there are great advantages, but there are also limitations to serology models."

Viboud: Yeah, absolutely. It's sort of that we're in the infancy of looking at this data and people are very excited about it, but we don't exactly know how to treat it. The human system is complex, but everybody responds slightly differently to a pathogen exposure. That's more difficult to pick up in a test—much more difficult than picking up the DNA [deoxyribonucleic acid] or RNA [ribonucleic acid] in a PCR test. Also, when you have a system, which is what we have now for COVID or flu, where people have been exposed multiple times in life, you don't really have a very clear signal where you didn't have any antibody before and suddenly you do if you've been exposed. You have a background of antibodies because you've had all of this history over time, and so you might have a little bit of a change, but not necessarily a very big change. People are now struggling a bit to interpret these complex serology data. That is true. But I think there's a lot of hope and also a lot of methodological development in terms of trying to understand these data and how we could use them for predictive approaches.

Barr: The effect of vaccination was another key area you examined. Will you describe the study designed to provide global, regional, and national estimates of target population sizes for COVID vaccination to inform country specific immunization strategies?

Viboud: This [inaudible] was relatively simple. It was mostly based on census data, like how many people you have in different age groups, and then also trying to use serology approaches that will tell you how many people have already been infected in the different age groups. We also try to get estimates specific for priority population groups such as healthcare workers, etc., which we knew would be prioritized for vaccination almost globally. It was mostly data from the census combined with some idea of how many infections there had been in the population already, so how the different age groups were immune, or not, to COVID-19 from natural infection.

Barr: You worked on China's prioritization in particular. Will you shed light on how the prioritizing of whom should receive the vaccine in China was devised, including some of the elements that were a part of making the decision and the types of analyses used?

Viboud: It was an extension of the same approach we used globally, just for China. I think we also wonder a little bit about the provinces in China. It's very big, obviously. The coverage rates are very high, very quickly. We were wondering how vaccination could be distributed spatially and were also thinking about different target age groups. Again, do we want to try to vaccinate or prioritize high transmitter groups, if we can find them, or groups with severe disease? We sort of always aimed towards the severe disease age groups. I feel that's what most countries ended up doing in the end, because it's such a strong severity profile with age with this disease, that it's hard to beat that with just trying to reduce transmission.

Barr: In China, especially in 2021, you were involved in a number of studies that looked at vaccine-induced immunity and what it would take to achieve that, as well as the role of vaccination and non-pharmaceutical measures being needed to keep the virus at bay. Can you talk a little bit about some of your work on those issues?

Viboud: The big thing with China is that they had been so good at controlling the outbreak that there was almost no natural immunity—because almost no one had seen the virus. If you start releasing things, you're going to have a very large outbreak. To prevent that, you really want to vaccinate as many people as you can. This is a huge country so that is s very, very difficult. The other issue is that by the time they got to thinking about doing this, the virus had evolved very high transmissibility. This is why we started with an R0 number [basic reproduction number] of secondary cases per primary case of around three for the initial strain of Wuhan strain. By Delta, we were at nine, and Omicron, maybe twelve. When the transmissibility is so high, it's very, very difficult to control and you need to vaccinate many, many people. You need to have immunity in almost the entire population, otherwise you're going to see resurgence. Basically, that modeling work suggested that you would need to keep NPIs [nonpharmaceutical interventions] for a long time, in addition to rolling out a vaccine, so that you could wait until vaccine coverage was very high and then you could relax your NPIs. Then you would have a strong immunity wall in your population and the virus wouldn't be able to really make too much damage. In practice, this couldn't be done. In honesty, it's very hard to keep very drastic NPIs for very long. That's part of it. NPIs were probably released too early and so they had a very big Omicron wave in China, but the modeling had suggested you can't try it because you really need to build a high level of immunity in your population. They just don't have anything if they haven't seen the virus, and so you need to ramp up coverage. An issue also specific to China was that there was a lot of hesitancy in older people—which is kind of the reverse of what we had in the U.S. and Europe—because of traditional medicine. The really vulnerable population actually was left vulnerable because they didn't have very high vaccine coverage.

Barr: I actually had a question about that for you. In your models, how do you account for different efficacy and durability and rates of different types of vaccines? Russia and China had their own vaccines; in the U.S. there were a couple of different types that had different levels.

Viboud: Yes, we can account for that in the model. It does mean more complexity. Some of the China models were very detailed and could handle that because they've used at least two vaccines in China. In the U.S., we struggled a bit in the beginning, because we had the Johnson & Johnson vaccine, which had quite low efficacy, and then the two mRNA, which sort of went head-to-head—but then Johnson & Johnson was dropped very quickly. Then we could just assume that both RNA vaccines were behaving the same and that's what we've been

using so far. We have just one type of vaccination in a way, but every time there's a new variant, you get slightly different estimates of VE [vaccine effectiveness]. There's also waning, etc., so we tune that accordingly.

Barr: When you're dealing with global studies and people get different kinds of vaccines all around the world, how do you account for all that heterogeneity? That must be very difficult.

Viboud: You really need different models for different countries, just because, first of all, population structures are very different. Their experience with COVID—so serological attack rate for natural infection—are very different. The NPIs are very different. Then, as you said, the type of vaccine and their effectiveness is different along with the coverage rate. What is really important is coverage times effectiveness, so what portion of your population is actually immune. Absolutely, there's a lot of heterogeneity—and not quite enough data to really fit country level models everywhere around the world.

Barr: There was a similar study that you did in China—but you also did a few in the United States—where you looked at different kinds of preventions, such as vaccines, non-pharmaceutical interventions, and testing. Can you speak about some of those models you were a part of and the public health outcome?

Viboud: This was part of a very big effort in the U.S. that started in December 2020, called the "U.S. COVID-19 Scenario Modeling Hub." It's to generate long-term projection of where the epidemic is going to go. Long-term means between three months to twelve months, under different scenarios of interventions, or different assumptions about new variants that may or may not arise, and how much of the population is going to pick up a vaccine. What's quite unique about this study is that it relies not on a single model. All of the studies that we talked about before were relying on a single model analysis, but this is using more than one model. It's the idea of crowdsourcing or combining across different models or different experts to try to beat uncertainty. You want to do this in periods of high uncertainty, such as pandemics. We've been going through 17 different rounds of projections, and four rounds for flu—because we expanded to flu more recently—working very closely with CDC on different vaccine strategies for primary shots, and now for boosters. But also, to try to help with planning, for instance, around the Omicron wave. People were worried about the hospital capacity. We published quite a few papers, and all of our data is publicly available online. I should also say that every single projection includes between six and twelve different models. That's really a lot of work from a lot of academic teams across the U.S.

Barr: In the United States, what are some of the reasons it was difficult to predict the number of cases and hospitalizations in the latter half of 2021, despite the fact that the models accurately showed the projection of surges and when they were going to occur?

Viboud: These were projections that we released just before the Delta variant wave. That was done in July 2021. And indeed, we did predict results, and we did predict which states were going to be the most affected—essentially, the states that had very low coverage by that point—so that was right, but we underestimated the magnitude of the Delta wave. When we look at different statistics and performance, this is the round where we did the worst of all of the rounds that we ever did. We think maybe the reasons are multifactorial, but the main one is that our estimates on vaccine effectiveness were too high. We only relied on estimates from the very early intervention trials, which were still using the Wuhan strain. People were vaccinated immediately before they were tested. Most people by that point in the Delta wave were several months into their first vaccination. Then we learned that the immune response is waning, and your antibody declines over time. We just didn't know that by the time we released the protections. The other thing that was quite important is that Delta is now well known to be the most severe variant we've ever seen. It was 30% more severe. Those data only became available later in the Delta wave, so we didn't have that. That affected our estimates of hospitalization and deaths. The last thing is

that this was also the first summer when everyone relaxed. There were a lot of behavior changes, and more contacts between people, which we have sort of tried to protect but perhaps not quite enough. All of these compounded to make our projections more optimistic than what happened. Our public health partners still found them very useful, especially knowing which states they had to concentrate on because they would be most affected.

Barr: What's it like to predict human behavior, because sometimes that's the hardest part of all? People went crazy that summer—they were so scared in March and then by July they were done with it.

Viboud: You're vaccinated and you're like, "Okay, I'm over it. I've been good for a year and a half." The behavior of the parties is very tricky. In infectious disease modeling, it's kind of easy for us because, as compared to climate, we don't have so many nonlinearities—susceptible people coming into contact with infectious ones. We've come to realize that we have more complexities because of behavior. In climate, behavior doesn't really affect things—it's physics. Actually, this was realized a while ago, during the Ebola outbreak in West Africa, because there are a lot of the models that were actually quite pessimistic because they didn't anticipate that people would just get scared about it and reduce their contacts. By that point, we knew behavior would become important, but we still don't quite know how to handle it. As you say, measuring behavior is difficult, but predicting is even worse. Most of the predictive models for behavior has very low explanatory power. I think that's a very good question, and still a nice area for future research.

Barr: Will you discuss your role in the multi-model study that looked at the impact of COVID vaccination of children ages 5-11 years on COVID-19 disease burden and resilience to new variants in the United States from November 2021-March 2022?

Viboud: This was also an output of the scenario modeling hub, so the same type of multimodal projection. We were working with an ACIP [Advisory Committee on Immunization Practices] group at CDC that does recommendations. They were thinking about whether to extend the COVID vaccine program to children five to 11 years old. By that point, fall of 2021, children were not vaccinated yet. There are many considerations that come into ACIP recommendation, but one of them is the population level benefits of a vaccine program, which is something that only models can really look at. We ran a set of scenarios for them, and we saw that you will get quite measurable benefits if you stand up a vaccination program. We also assumed in some of the scenarios that there will be a new variant coming in. This was the Omicron period, but we didn't know that Omicron would be there. A hypothetical variant didn't look like Omicron, but that was suggesting that if you have a really big stress on your system, you really want to have as much immunity as possible, and so vaccinating children would be a good idea.

Barr: In 2021, you were part of a team that developed a computational model that looked at the effect of school closings in mitigating the transmission of COVID-19 in many countries, including Finland, Latvia, Ireland, Slovakia, and the United States. What were the factors in putting together such a varied model, and what was ultimately learned?

Viboud: Yeah, so I'm completely blanking on this study to be honest. We should probably move on to the following one.

Barr: Will you discuss your role in reanalyzing incubation-period and serial-interval data describing the transmission of the Delta and Omicron variants from the Netherlands at the end of December 2021 and how that was important to estimating the reproduction advantage of the Omicron variant?

Viboud: Yes, there are some parameters that are very important for disease modeling and just for thinking about transmission, and the serial interval is one of them. It's the time interval between a case and a transmission in another case. If you make errors or if your estimate of the serial interval is wrong, then all of your projections about what the epidemic is going to do and how to control it are going to be wrong. Because the SARS-CoV-2 virus evolves a lot, it is completely possible that the serial interval also changes over time. This study was led by Princeton. They looked at contact tracing data from the Netherlands. You need these chains of transmission, one case to the next, to look at serial interval. There's a difference in the serial interval of Omicron versus Delta. Again, that is important, because then that affects the assumption that you're going to make about how to control the outbreak.

Barr: Will you talk about a study that you were part of that looked at the behavioral factors in the transmission heterogeneity of COVID-19 in Costa Rican households, and how your team applied a chain binomial model to the serologic data collected to account for exogenous community infection risk and potential multi-generational transmission within the household?

Viboud: Yeah. This was an interesting study that was done in Costa Rica, where historically there have actually been good collaborations with NIH on the HPV [human papillomavirus] vaccine. They've built on that to look at SARS-CoV-2. It's a pretty traditional study design, where as soon as there is a case in a household, and you enroll everyone in the household, you try to monitor transmission, which is what we call a case ascertained transmission study. Then transmission was assessed by serology. There were many, many questions about what behavioral precautions people took during the outbreak. And it also discovered quite a few waves. The main thing was that taking precautions within the household works. In particular, if the primary case, the index case, wears a mask, that does limit transmission to other people in a household. Sometimes we can think if someone has it in the household, everyone will get it, and it's kind of a given. But no, that is not true. You can actually prevent transmission by masking, even within a household where contacts are very, very dense. There were some differences in the number of bedrooms and in age groups. It's better if you can sleep in separate bedrooms, obviously, when there's COVID. The main thing was the good news about the effect of behavioral interventions. That was also a pre-vaccination study.

Barr: Will you speak a little bit about why the chain binomial model was chosen to answer this particular question, and, in general, how you and others in your field decide to use what kinds of models on what kind of analyses to answer particular types of question?

Viboud: The negative binomial model has been used a lot for household transmission since the 1960s. It's a model that is only based on the final size of the outbreak. You just want to know, after the outbreak is over, how many people have been infected in your household. It's very nice because you don't need to follow infection over time very closely. You don't need to know every day who was infected and who was not, you just need to know who's been infected at the end of the outbreak. Serology just tells you whether someone had it or not. It doesn't tell you when they had it, so it's kind of a crude estimate in a way. If you do serology, that's the only model you could use. But it is a neat model because it takes into account the dynamics and potential chance of infection within the household, so that when you see a case, it actually could be infected by another one or it could infect another one. There are many different directions that pairs of transmission can go through. The other approach that people have used is just a logistic regression, but that doesn't make all of these assumptions about the directionality of transmission. It's less of a dynamic model. It really depends on the type of data. This is a good model for, in a way, small data sets—household level. The models for something much bigger, like a city, a state, or a country—like we did for [inaudible] or like what we did for China—are using

completely different approaches. They are using a compartmental dynamic transmission model, where you put a lot of people in the same group and just assume they react the same, because you can't track everyone, there's just too many. There's quite a big set of tools in modeling that depend mostly on the type of data and somewhat on the type of questions.

Barr: Will you share your role in the prospective cohort study that looked at the Omicron variant in South Africa and how it shed light on population immunity and circulation of future variants within communities that have already been exposed to the virus to some degree or have been vaccinated? That's a very different scenario than some of the earlier studies.

Viboud: This is South Africa. It's a completely different setting, right? We've worked on a couple of studies. This was led by Kaiyuan Sun in our group. They're from very detailed cohorts. They're household as well, but in contrast to Costa Rica, this one tested people twice a week—no matter whether they're symptomatic or not—from PCR, and then they also do regular serology. You have the complete timing of infections, as well as the immune response, with serology and antibodies. There's also good surveillance in the community. You can really track the number of infections per person, and also what type of variants they've been exposed to. I guess the question there was that there had been a lot of work on the protective effect of the vaccine, but not so much on natural immunity. In South Africa, when we looked at that data around the time of the Delta wave, they had very little vaccine coverage—around 6%—so it's really mostly natural infection. We saw that natural infection provided very long-lasting immunity—up to eight months, we think—against infection, so what we call sterilizing immunity. It's not just against severe disease—you're not going to get infected, and so you're not going to transmit. That's stronger than vaccination. Even against the Beta variant, which some had said was antigenically different, but it was probably not that antigenically different—and also against Delta. When we got to see Omicron, it was completely different, because then we could see people get reinfected even through this very strong natural immunity, because it was a clearly antigenically different variant. We got about 60% reinfections there. By the end of the Omicron wave, people had been infected, on average, maybe 1.5 times naturally in the study, which is a lot compared to anyone anywhere else—for instance, in the U.S. or Europe.

Barr: Will you introduce the EPIFORGE checklist and how that has helped in the COVID-19 pandemic and may be used in future outbreaks?

Viboud: This is kind of a practical study. There's a lot of interest in modeling and with COVID, it's just been duplicated, triplicated, etc. There's so much modeling work. We realize that the way people report their models is not very standard. We have reporting guidelines for clinical trials and for a lot of study types, but not for models. EPIFORGE is trying to set up common guidelines for people to report their models, including specifying what the calibration data is, what you are fitting to, whether you've done sensitivity analyses, how you calculate in confidence intervals, and whether you're trying to do forecasts, short-term projections, or longer-term projections. We came up with a list of maybe 15 to 20 checkmarks and other items that you would need to fill out if you wanted to report your model results, and we hope that it's been picked up by the community. This is relatively new—the paper came out during COVID-19—but we hope this is going to help standardize the field a bit. Another big question in the field is making the code and the data available publicly so that people can use your model.

Barr: How did your experience with other epidemics such as influenza prepare you for COVID-19? What have you learned from COVID-19? Models were just so important with this outbreak.

Viboud: I think a lot of what we had built with flu was useful. Some of the models, for instance. It is directly transmitted respiratory virus, so a lot of the COVID models built on the flu models—all of the excess mortality approaches that have been developed for flu and pandemic flu. There had been a lot of work during the 2009 flu pandemic to try to think of the minimum dataset we needed in order to look at, for instance, a case fatality rate early in the pandemic. There were quite a few things we knew. At the same time, we could even argue that this knowledge that we had from flu was detrimental in some cases, because in flu, you do not get a lot of transmission from asymptomatic [people]. That was also the case for SARS-CoV-1, the first SARS outbreak. But it was completely different for SARS-CoV-2, and we were not quite prepared for that. We were surprised by it. The only thing that we didn't anticipate at all was the rapid mutation. This idea we have from pandemic flu is that if a new virus emerged in a population where everyone's susceptible, it's going to take a while, probably years, before it really evolves. This SARS-CoV-2 virus evolved very, very quickly with increased transmissibility and completely new antigenic shape or features. Those were things we had not anticipated at all from our past experience with pandemics. I think we learned a lot from that.

Barr: Do you have any concerns about data used for models over the course of the pandemic? In some cases, there is underreporting or a lag in reporting, and there are cases that were attributed to COVID that maybe weren't. What were your thoughts on that in terms of your models?

Viboud: Yeah, it's a really great question. And it's something that people know, and there's a lot of issues there. But sometimes as a modeler, you don't spend a lot of time thinking about that. You think more about the ingredients in your model or how many components there are. There's been a lot of effort from groups who are more statistically oriented and do not do the transmission modeling—they just do the data path to actually clean the data, and try take into account, for instance, reporting delays. They maintain repositories of data and clean what people have been using. The prep work is not done directly by the modelers. Instead, it is done by statisticians who are interested in these models, but it's sort of a pre-cleaning step. That's very useful. A lot of these tools are being developed now. It is very important, especially in a pandemic, where lags are going to be long early on, but they're also going to change over time and that's going to affect your estimates. I would also say that, because there's so much uncertainty there and issues with handling the data, having different models is really good, because everyone handles the data part differently. You average across all of this uncertainty.

Barr: What have you found to be most exhilarating about your COVID-19 work? You've worked with a bunch of different countries and issues.

Viboud: It's been a great opportunity just to learn about a new disease. I'm going to take a very scientific approach because of course, this was a terrible thing. From a scientific perspective, you get to see the emergence of a new virus and invasion and how the population will act. We have relatively good data, right? We complain about data a lot, but this is unusual—the level of information we had about these viruses—and there was so much sequencing. You get to learn all of that. On the immunology side, we also learned a lot. We get these new vaccines developed, but we also learn how someone completely naive gets to an immunity to this virus, and the virus keeps changing. We know this happened with flu. We just never get to see it. Scientifically, it's been really great—in particular, the global community coming together on Twitter, but also online and on Zoom calls, to try to work through problems. This has been a really good thing—but everyone's very, very tired now, of all of this.

Barr: Do you continue to be part of many COVID-19 projects?

Viboud: Quite a few. But we're sort of all hoping to move back to our other stuff. I mean, we know COVID-19 Is there to stay. If you work on respiratory viruses, you had to add that to your usual panel. At least the emergency part is over, which is good.

Barr: In addition to being a scientist, you're also a person who has been living through the pandemic. What have been some opportunities and challenges that COVID-19 has presented for you, and how have you coped with it as an individual?

Viboud: I think it's allowed me to work with international collaborators that I had in the past and strengthened those collaborations, in particular with China and South Africa, so that's been very good. The challenges are similar challenges to what everyone has experienced—that even though we were on Zoom a lot, and we could communicate, we could not see each other. A lot of scientific collaboration comes from regular contact. But mostly, it was the additional opportunities [it presented] because of the opportunity to learn about a completely new virus.

Barr: Is there anything else that you would like to share about your COVID-19 work and experiences?

Viboud: No, I think that's it. We went through a lot.

Barr: We definitely did. Thank you so much for all your work and effort. I wish you and everyone you work with only the best. Stay safe; it's going around again.

Viboud: I know, it's coming back up! Yeah, it's definitely here to stay. Thanks a lot for putting all these questions together.