

Monitoring Chikungunya with Big Data

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

[Sarah Gregory] Hi, everyone out there. I'm Sarah Gregory, and today I'm talking with Dr. Joacim Rocklöv, who is calling in from Sweden. Dr. Rocklöv is a professor at Department of Public Health and Clinical Medicine at Umeå University. We'll be discussing his article about using big data to monitor a 2017 outbreak of Chikungunya in Europe. Welcome, Dr. Rocklöv.

[Joacim Rocklöv] Thank you. Thank you for inviting me to today's show.

[Sarah Gregory] Okay, let's just start with very basics: What exactly is Chikungunya?

[Joacim Rocklöv] Chikungunya is an infectious disease. It's a viral disease that's transmitted by mosquito vectors. It's been quite rare for a long time, but it's kind of on the upsurge, so it's becoming more frequent. Particularly interesting is that it's spread by one of the mosquitoes that's competent to spread it is *Aedes albopictus*, which is a very invasive mosquito, sort of expanding its territory from traditional, you know, tropical, subtropical areas to temperate zones. So, Europe and the U.S. is nowadays quite a lot of *albopictus*, *Aedes albopictus*, and their great...their very great capacity at vector competence is high for transmitting Chikungunya virus.

[Sarah Gregory] You began tracking Chikungunya in 2017. Why that date and why in Europe?

[Joacim Rocklöv] Outbreaks in Europe, so far, are quite rare in Chikungunya. It's kind of on the edge of where transmission could occur, because, you know, transmission zones are partly dictated by where the mosquitoes are, but also the climate. It's very important for the vector to be competent...sort of, in replicating virus and also surviving enough long to transmit the virus to humans. So, this is the second big outbreak of Chikungunya in Europe. And we were just basically ready this time to track it and to investigate the spread in almost real time.

[Sarah Gregory] Your study is about using big data to track the spread of Chikungunya. Now, tell us what big data is.

[Joacim Rocklöv] Big data is big...[laughter]...it can mean a lot of things. Particularly, it's untraditional, you know, big data in terms of the public health and the medical sciences, you know. Because of the wealth of data and the speed and the volume and the velocity of data, and that...also the variety of data available nowadays, in terms of when you put everything together and when you have data in almost, you know, almost near-real time, if you want to make something out of that, you need to have special methods, and that would, you know, kind of qualify as big data. But it could mean a lot of different things.

[Sarah Gregory] Well, in this case. So, could you give us some examples of, you said, a variety of different datas that you're bringing together into big data. Could you give us some examples of what those might be.

[Joacim Rocklöv] Yeah. So, for example, we brought together a variety of different climatological datasets, and then we used specific ones that are important for...that we now have, you know, biological possibility in terms of the vector and the vectorial capacity. We also put together flight data from, you know, all international global...passenger travel from endemic zones to the outbreak zones in Europe. We also mined Twitter data, and that's the biggest, you

know, the fastest and the most voluminous part of the analysis. We didn't use most of the Twitter data, we didn't care about the messages, we just cared about trying to see, you know, trying to use the year curves of aggregated users to understand how people were moving around. As it has shown in, you know, another study we've done, predictive of, you know, the propagation of outbreaks locally.

[Sarah Gregory] So, apparently other sectors of society are doing things with big data that public health is not doing. What are some of those things?

[Joacim Rocklöv] I mean, I think the most advanced sector is probably industry, with documentation of processes in industry production, etc. where you can have a lot of sensing, sensors, on your, you know...I don't know what you call it...but industrial production lines, etc., sensing how the quality, the performance, everything, and then mathematically, basically, calibrate the process. So, that would be...that's a highly advanced area. Public health, there's a lot of talking about big data and, nowadays, also artificial intelligence, but there's less actually being done.

[Sarah Gregory] So, okay, so less is being done—so, why should big data be important to public health?

[Joacim Rocklöv] I mean, there's a lot of information out there that could be used. If you want to intervene and understand, for example, in this case, an infectious disease, how it's going to potentially or most likely spread, which is important for prevention and control, then a lot of these sources can actually help you make a better decision and, you know, informed decision. But, because the data sources and the method, there's so much data, you would need to use sophisticated methods, processing and filtering of that information, for it to have any value to your decision making.

[Sarah Gregory] So, tell us about your study.

[Joacim Rocklöv] Well, basically, what we did is we used biological knowledge about vectors and the relationship between the vector and climate conditions and seasonality in their capacity to basically reproduce Chikungunya virus. So, that's kind of, you know, informed from the literature and we predicted the mosquitoes' ability to transmit virus in...all across Europe. Then we sort of sifted out signals of incoming virus from these air travelers from...in areas where there's actually active transmission of Chikungunya virus. So, that's one part of analysis trying to, you know, understand where from and when, kind of side.

The second part of the analysis is that we also extracted when the outbreak actually had occurred in Europe, in Italy and in France. We also extracted these mobility data from the Twitter about users to investigate how the onward spread would look, potentially, and most likely look like, both locally and at the regional level for...as a, sort of, to showcase for decision makers that this could be useful. And then we combined the vector capacity signature of the vectors, in relation to climate, with the mobility for which you see high-risk areas, where the mosquitoes...where the virus was most likely to be onward spread, and also, to look at, in detail, on the seasonality for these localities, so you can see, you know, how long could the outbreak be sustained before the season sort of forces it to stop.

[Sarah Gregory] As you kind of mentioned before, your primary sources of data focused on climate, air travel, and internet and social media activity. Why did you choose these sources

specifically, and how is each one useful, particularly, in tracking in the spread of mosquito-borne infection?

[Joacim Rocklöv] Yeah, so, the air traveling is, from a...from a European sense, given that Chikungunya is not a disease that's, you know, endemic or frequently or continuously transmitted in...you know, locally. The air traveling is important, in combination with other, you know, transmission zones to understand the importation rates of the virus into...to areas where there's no transmission.

And then you need to know something about the vectors, if the vectors are abundant in their seasonal activity and their capacities. And the climate directs a lot of the actual disease and all, and the pathogens, basically, of the vectors—how they sort of reproduce and how good they are at, you know, replicating virus. So, that's another thing, important part. And then we used this novel source of the Twitter mobility proxy data for...for trying to understand something that's not usually available for us—the human potential movements at the local and regional scale, which is important, in combination with known transmission areas. It's important, of course, how many people go from these areas to other areas. And would it be potentially so that Chikungunya could be spread in these areas where people are going to from the *epi foci*.

[Sarah Gregory] How did you get access to these sources of data? Are they public?

[Joacim Rocklöv] Everything is not public. The public part is Twitter, which is from the Twitter, Twitter open APIs. It's a Twitter server. It's not...it doesn't include the complete...all tweets, basically, but it includes a reasonable sample. And then, the climate data is freely available, publicly available. And then the vector abundance...or the vector presence data is also available through communication with, for example, your CDC or the European Centre for Disease Control. Air traveling data is not freely available at this stage, so that's a source which you need to negotiate or collaborate or buy your way into.

[Sarah Gregory] What about the climate data?

[Joacim Rocklöv] Climate data is freely available. There is a lot of climate databases and the development in the area is kind of moving towards making more and more data available. Sometimes when you want to get into highly resolved satellite information, it could be...it could be more difficult and you need to pay for your subscription or data, then. But a lot of the climate data is also available for free, from NOAA, for example.

[Sarah Gregory] Maybe this is not something you actually know the answer to, but you collected this Twitter data, so as a Twitter user myself, should I be concerned about my privacy when I post something about some private ailment?

[Joacim Rocklöv] I think for all data sources, with social media, when we give out our geocode, for example, I don't know if you do, I don't...I've done that myself, actually. I think we should think about the consequences of that. In our analysis, I don't think you should be concerned, 'cause we aggregated the analysis. So, we didn't look at individuals, we just looked at, you know, flows of people from one area to another. But, yeah, I mean, there is good and bad sites of all these data sources and social media, and there would probably need to be a bit more regulation. The GDPR is a good development. But, at the same time, we should also know, using this for good stuff, for things that are, you know, valuable for us as people in society and for decision makers and then public health, of course.

[Sarah Gregory] Okay, so going back a little bit, we'll just go back to your study. What were the results?

[Joacim Rocklöv] The results was, basically, that we could highlight risk areas for the onward transmission, and worked out pretty well. We could also actually come to the conclusion that the transmission wouldn't be in Europe at the...in the current climate situation. We didn't think the outbreak could be sustained over the winter season, as well. So, there are two important results for policymaking, which means it's going to be controlled naturally, in a sense.

But these things could actually be useful for...for...for investigating spread in outbreak situations. And that's also something that we've seen. I've been in situations with those studies in Indonesia, where we looked at the local spread within the city of Yogyakarta, and the mobility, the local mobility rate, also from Twitter, is...is, you know, it's very highly predictive of the propagation and the spread and, you know, the where and when of outbreaks in the city at a very fine scale.

[Sarah Gregory] What was the most challenging part of this study? Were there any unexpected ones?

[Joacim Rocklöv] No, not really. I mean it's always, in this kind of outbreak situations, when you do a study, it's always difficult, because everything constantly changes. And we started basically, you know, in the outbreak. And so, that's pretty challenging because you have to continuously update everything. But otherwise...yeah, it's also a bit challenging still with this kind of big data sources and handling and management and presentation. How do you...how should we, you know, filter out these wealth of information to something that's efficient enough to use as...as an indicator for the spread of...for, you know, health policy...you know, something that's important for public health?

[Sarah Gregory] What are the next steps for public health officials and big data? What do you think should be done next?

[Joacim Rocklöv] Well, I think that it would be important to do case studies, just like the one we've done, and show how big data is actually and can actually be used in situations like this. And talk less about big data and machine learning and artificial intelligence, and actually show proof of principle. And then, I think there's also like this capacity building component to it, where it's actually...because it's not the traditional way of working, it's not the traditional data source. I think there's actually the need for strengthening our knowledge about how to use it—how the decision-makers' and peoples' knowledge about the capacities to use this type of data sources and information and outputs from the analysis.

[Sarah Gregory] Are there future studies you would like to see done on this topic?

[Joacim Rocklöv] I'd like to see a lot of future studies on this topic. I think it's, you know, outreach is really promising. Of course, all data is not good. And there is still like, you know, like in classic studies, you need to have good data, and you're probably going to throw away a lot of the information you have, in the end. But there is also a lot of potential that shouldn't be...go lost in this rich sources of data. Just takes a lot of sophisticated analysis to make use of it, which is also, I think, there is a little bit need of a capacity building in the community to actually better harness this type of data sources.

[Sarah Gregory] Tell us about your job. I know you're a professor. What do you teach and what do you like about it?

[Joacim Rocklöv] I like my...I teach epidemiology and statistical mathematical methods in medicine or epidemiology and public health. I like the research that we're doing because it's cross-disciplinary, it sort of crosses the boundaries of a lot of different disciplines, such as medicine, infectious diseases, environmental sciences, climate sciences, entomol...medical entomology, and tries to sort of bring it all together to make sense, in terms of understanding disease and disease spread and propagation, risk of disease, from a system point of view. It's really fun. And, you know, what I like most about my work is to use the creative process in trying to come up with new, interesting studies that we can do.

[Sarah Gregory] Did you, as an undergraduate—is this what you wanted to be doing or how did you end up here?

[Joacim Rocklöv] Not at all—I had no clue. But I'm really happy with how it's all developed, somehow, by chance. Random.

[Sarah Gregory] I assume you have a Twitter account. Do you use it to post about your work results or do you post about your sore throats?

[Joacim Rocklöv] More work, more professional Twitter. Sorry. I do it once in a while, but I'm not continuously Twittering enough. Well, it takes...

[Sarah Gregory] Yes.

[Joacim Rocklöv] It eats up too much of my time.

[Sarah] So, but I assume you will tweet this podcast when it's live?

[Joacim Rocklöv] Yes. Absolutely.

[Sarah Gregory] Yeah? Good! Thank you so much for taking the time to talk with me today, Dr. Rocklöv.

[Joacim Rocklöv] Thanks very much.

[Sarah Gregory] And thank you, listeners, for joining me. You can read the June 2019 article, Using Big Data to Monitor the Introduction and Spread of Chikungunya, Europe, 2017, 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.