

[COVID Information Commons \(CIC\) Research Lightning Talk](#)

Transcript of a Presentation by Jennifer Surtees (University at Buffalo), January 31, 2023



Title: [Center for Ecosystems Data Integration and Pandemic Early Warning Systems](#)

[Jennifer Surtees CIC Database Profile](#)

NSF Award #: [2200173](#)

[YouTube Recording with Slides](#)

[January 2023 CIC Webinar Information](#)

Transcript Editor: Lauren Close

Transcript

Slide 1

Good afternoon, everybody thanks for being here. I am excited to talk a little bit about what we are doing. This is our PIPP for pandemic prevention project funded by the NSF that Mark [Lurie] talked about earlier. So this is another one of this mechanism. My name is Jennifer Surtees and I am at the University at Buffalo.

Slide 2

I just wanted to start by sharing with you what our grand challenge is for pandemic prevention. So basically we are the Center for Ecosystems Data Integrations (CEDI) for Pandemic Early Warning Systems. Our Grand Challenge is to basically develop a number of streams of data at different levels. To develop this - that we can integrate to develop an early warning system for outbreaks and pandemics. So the idea is to integrate multiple diverse databases that increase the viral - I'm sorry -

Slide 3

The viral ecosystem which is derived from wastewater samples and patient samples. The human ecosystem based on big data, GPS data, cellphone data, as well as community-level ecosystem information, based on talking to the community and members different members of our community. We see this development of data and integration into an early warning system combined with communication of our findings, solicitation of feedback from our community members, and generating information from our community members to integrate into our early warning system. As this feedback loop mechanism for generating this more robust ecosystem or early warning system. So we have as a goal to develop this early warning system but also as an

equally important goal is to engage with our communities and build real partnerships, not only for the generation of data for our early warning system, but also as a mechanism to have real partnerships and a trust relationship with our various community members. It doesn't matter how good our early warning system is if, when we do issue warnings, the community doesn't respond. So building that trust back is really critical.

Slide 4

I'm not going to go through this entire slide, I just want to share with you the fact that this is a highly interdisciplinary group and a growing group from across the University, across the disciplines, ranging from genomics to mathematical modeling machine learning, engineering, virology, history, disability studies, chemistry, etc. This is a terrific team.

Slide 5

Outside the university we have partnerships with different health care systems. New York State - we're part of a New York state whole genome sequencing consortium at the moment - and the Buffalo and Erie County Water Supply.

Slide 6

We have this fantastic network to help us generate and implement - integrate and implement the data that we're talking about. We're trying to use our region in Western New York - so again I'm in Buffalo and we're in Erie County - trying to use our region as a community lab and to generate data based about our microbial ecosystem. What's happening in terms of the viruses and microbes that are present through both wastewater, which I'll talk about and individual patient samples. I'll talk more about that in a few minutes. Also understanding our human ecosystem. Our community relationship and capacity building. We, through a community advisory board, the development of surveys and focus groups, again which I will talk about in a little bit, but also data mining and GPS monitoring and GPS building and network building for for the purposes of modeling. Developing these different types of data sets at the microbial system, at the human level, and then putting them into this predictive modeling Hub based on mathematical modeling and machine learning. We then have that output go out to our community members, our stakeholders, as well as to going but feeding back into our analysis of our ecosystems. Really critical to all of this is the communication to our stakeholders - to our community members. Communication of the science, communication of risk. What level of risk are we at?

Slide 7

What underlies a lot of what we do is sequencing and so I just very quickly want to orient you to SARS-CoV2 sequencing and then beyond to other microbial cleansing approaches. I just want to orient you to what we're looking at when we look at these phylogenetic trees. Basically, we start off with an original SARS-CoV2, the Wuhan Virus, if you will, and then over time as the virus spreads, it accumulates mutations. We get these different lineages and variants of concern ultimately. As we get changes, we get the branching off of this phylogenetic tree. So in this particular case, we start with this blue sequence. But then on this clade it's accumulated another new mutation - this orange circle - and then another mutation, this green circle. So these two

individuals, A and B, have the same lineage. It's this combination of mutations. Lineage C is similar, but instead of having this green circle it's got this red circle. So while it's closely related, it is distinct from lineages A and B. Similarly, we have this other branch point where we have related lineages but with some differences.

Slide 8

So what we've done here in western New York is - and people have done this all over the world, but this is looking at western New York - this is our phylogenetic analysis of what we've seen throughout the pandemic. We've been able to build these phylogenetic trees to really assess what's present in our communities.

Slide 9

What we've also done - and this ties into our modeling - what we've also been able to do is look more closely at the different lineages to do a cluster analysis. So by taking in this case all of the alpha variants that we have identified in our region and clustering them based on the similarity of the sequences using those phylogenetic trees, we see distinct clusters. So the more yellow it is, the more similar the the linear the alpha lineages actually are. We see these distinct clusters labeled one, two, three, and four, which are all Alpha variants. Then we have background mutations which make them distinct. By looking at these these differences, it suggests that there are multiple introductions into our community well as community spreads. So different versions have been introduced and once they're introduced they spread to other people. This is just shown you can see this in these phylogenetic trees. These arrows are pointing at actually unique mutations that are actually only ever seen in western New York which is kind of cool.

The other thing and we can actually use the spread - the tracking the spread of SARS-CoV2 to actually do agent-based modeling, which I'll talk about in a second, to actually understand how these different versions of SARS-CoV2 move around in our community and model that movement based on some of the human data that we will also integrate.

Slide 10

One of the things that we did notice - so this is actually looking at the different lineages of SARS-CoV2 across New York state. So these are the different regions of it within New York state and and the different panels represent the lineages over time, in this case in 2020, in these different regions. If you just look at the color patterns you can see that these - and actually we also included Ontario because it's nearby to Buffalo -but if you just look at all of the different patterns, the color patterns in these different regions, you can see they're really significant differences in the patterns of which version of SARS-CoV2 is present, even in regions that are very close together. And this was even more evident when we look in our 2021 [data]. I'm not going to show [that] but we this is also true in 2022.

Slide 11

We think we can use this distribution and variation in distribution over time as a way of modeling the spread of SARS-CoV2 throughout the state as well as within Erie County over

time. So this is just again showing a time lapse snapshot of the Alpha variant in New York state where it was localized and when with the more yellow being the more more prevalent. You can see it starting out in downstate and eastern New York and then gradually spreading. The timing of this variant is distinct depending on what part of New York you're in. This is just actually showing - this is just a .gif showing the Alpha variant moving in around New York state over time. This is just the New York variant to compare. You can see differential migration and differential movement of these two lineages over time.

By incorporating things like traffic patterns and developing agent-based settling, we're doing network analysis across New York state and within Erie County to understand the movement of the virus through time.

Slide 12

The other thing that we've been doing and this is work with our engineering collaborators Yin Yin Yi and Ian Bradley who, over the course of the pandemic, were tracking levels of RNA in our wastewater. This is a cartoon of wastewater - collecting wastewater data. This is our wastewater and then these are the SARS-CoV2 viruses that are present in our wastewater. Over the course of the pandemic, they looked at levels of RNA through as a function of time. Then about a year and a half ago we actually developed techniques for doing sequencing of SARS-CoV2 in the wastewater as well.

Slide 13

This is showing patient samples, individual samples, and again the colors are different lineages of SARS-CoV2. This is as a function of time on the x-axis. When we look at four different sewer sheds we can see that the type - the virus type itself - is similar among the sewer sheds although there are some differences. The patterns are actually different than what we see in the patient samples. So the wastewater samples is actually a pooled sample of everything in our community and so it's really interesting to look at the differences between a patient and the wastewater samples, patient samples being a little bit more biased in terms of just who's actually getting sequenced.

The wastewater can be used to monitor for different known and novel pathogens. So this is what we're actually doing now. In order to to really do this in a meaningful way we need community cooperation and trust.

Slide 14

Now I'm just showing you the same diagram, but here I've added in a whole bunch more squiggles and triangles to indicate that we're not just sequencing COVID anymore - we're sequencing all of the RNA that's present in the wastewater after depleting for mRNA and rRNA so that we can enrich for some of the microbial and particularly viral RNAs. The idea here is to develop a longitudinal picture of the viral microbial ecosystem over time so that we can have get a sense of what's normally there. If it's disrupted we can notice and then start looking at looking n more detail at what might be present. This is a way of detecting not only known pathogens but potentially an unknown pathogen as well. One of the other things that we're really excited to do

is to incorporate mass spectrometry to detect pharmaceuticals in the wastewater as well. It looks like the uptick in ibuprofen and that sort of thing actually precedes this an uptick or a spike in cases.

Slide 15

We're also looking at the community level to to really try to understand what biomarkers - what key questions do we need to ask our community partners in order to feed in modeling and anticipate an outbreak. We're collecting oral histories, we're doing focus groups and interviews with different communities across western New York. This involves a discussion of community concerns and interests, including the ethical issues that were mentioned earlier. We have a strong Community Advisory Board as well to help advise and to help us with our communication. Again, this data gets fed back into our models.

Slide 16

I also mentioned the importance of community engagement. We have actually a really robust track record and set of partners to do community engagement both in our K-12 schools and within our community more broadly speaking. These are just some images of of some of the community engagement events and activities that we've done over the past seven or eight years.

Slide 17

We're leveraging those partners now as well as a partnership with the Buffalo Museum of Science to do a lot of community partnership building, hands on type research.

Slide 18

This is one of our first sort of demonstrations for community engagement at the Museum of Science. We have just have this demo of what wastewater actually is and how we collect all of the those samples from within a neighborhood in order to get a big picture - a broad picture of what's going on at the viral level and the microbial level in our communities. [It shows] how we use that to predict and hopefully prevent pandemics. So I will stop there.