UC GIS Week
Tuesday, November 16, 2021
1 PM – 2 PM
Public Health-1

Sessions:

Spatial distribution of extended-spectrum beta-lactamase-producing Escherichia coli isolated from patients with community-onset bacteriuria

Speaker: Eva Raphael, UC San Francisco

Abstract:

Introduction: The prevalence of infections caused by extended-spectrum beta-lactamase producing Escherichia coli (ESBL-E. coli) is increasing worldwide. There is growing evidence that community-onset bacteriuria caused by ESBL-E. coli may occur as outbreaks. With data from a public healthcare system in San Francisco, we found increasing prevalence of ESBL-E. coli causing community-onset bacteriuria over 6 years. We examined spatial distribution of community-onset ESBL-E. coli bacteriuria to identify such outbreaks.

Methods: We collected electronic health record data on patients with E. coli

isolated from urine cultures in a public healthcare system from January 2014 to March 2020, including geocoded addresses. We identified community-onset bacteriuria as bacteriuria diagnosed < 48 hours after hospital admission or in individuals seen in primary care or emergency services with no hospitalizations in the past 90 days (N=7,751). We mapped all cases of ESBL-E. coli bacteriuria. We performed global and local Moran's I to identify significant proximity between cases of ESBL-E. coli bacteriuria and to identify spatial clusters and places with unexpectedly high ESBL-E. coli incidence, compared to the incidence of non-ESBL-E. coli.

Results: Patients with community-onset bacteriuria were predominantly women (86%) and mean age was 46 years. The study population was multiethnic, with 42% Latina, 19% Asian, 16% White, and 13% Black patients. Preliminary results did not show spatial clusters of ESBL-E. coli compared to non-ESBL-E. coli.
Conclusion: Preliminary maps show similar spatial distribution of community-onset ESBL-E. coli and non-ESBL-E. coli bacteriuria and no spatial clusters of ESBL-E. coli. We will be stratifying analyses by race and ethnicity as prior studies have found differential ESBL-E. coli burden in specific racial and ethnic groups.

Transcript: Video Timestamp: 00:00.11

of drug resistant, a special type of drug resistant E. coli isolated from patients with community onset bacteria.

So, first I want to kind of point to the crowd. I don't know how familiar you may be with drug resistance, but it's definitely a rising issue in the US. A couple years ago, the CDC came out with a new report and they realized that they had under-accounted drug resistant infections the last six years. And so, had updated it and just saw that there was definitely a very concerning rise in drug resistant infections. And so, here's some numbers. So, each year what they found is that drug resistant bacteria and fungi cause many infections, almost 3 million infections per year and more than 35,000 deaths in the US.

So I'm going to focus on urinary tract infections. My team and I are focusing on it as part of a paradigm that could be potentially applied to other infections. So if you have not had a UTI, here's the definition. So it's a bladder infection, symptoms are

the following. So having urinary urgency, pain with urination, usually caused by gram-negative bacteria, although there are some UTIs that can be caused by gram-positive bacteria. And the majority are caused by E. coli or Ischia or Yeschia coli, which is the bacteria that I'll be focusing on during this presentation. And I want to give a definition of ESBL. So it was in the title. So extended spectrum beta-lactamase. So that type of enzyme confers resistance to both ciftazidium and cifotaxine, which are cifalosporins if you're interested in the kind of classes of antibiotics and inhibited by plavidonic acid. And this one is of particular concern because it severely limits the number of types of antibiotics that we can use to treat an infection. So there are risk factors for UTIs and there are risk factors for drug resistant UTIs.

So previous antibiotic use, which is not to be expected, is associated with having a drug resistant UTI, having a recurrent UTI. But there are new risk factors that have been identified in the last decade or so that includes both travel, having companion animals, and food in particular poultry.

And ESBL producing E. coli, and I apologize if there's some noise outside.

So ESBL E. coli have been found in both food animals and environmental sources.

We will later be talking about secret sites. So there are kind of strains of or species genotypes are kind of like a specific fingerprint of E. coli. And so we can see that multiple genotypes have been found in multiple sources, including water sources, dogs, cats, and horses, retail chicken, house flies, and vegetables.

And one piece that is well, not so new, but there have been spatial clusters of ESBL producing introbacteriesia, which are the family under which E. coli fits under by neighborhoods. So I'm presenting here a study that was done in Colombia where they sampled isolates from

samples from two hospitals. And they did find some concentrations of ESBL producing introbacteriesia in various communes.

Here in the city, so in San Francisco, where our study, our own study is based, there have been clusters of another type of drug resistant infections. So methicillin resistant staph aureus, which can cause severe skin infections and even sepsis. And here what they found is spatial clusters of these community onset MRIC

infections, in particular the specific genotype called USA 300. So we were wondering, well, are we going to see the same thing in San Francisco with our samples?

And these are some preliminary findings that our work is based on. So what we found was that over the last six years, so from 2014 to 2020, the prevalence of community onset bacteria, so your bacteria in the urine that includes UTI and asymptomatic bacteria caused by ESBL, actually increased both in hospital settings and in community settings. So people who had not been hospitalized in last 90 days.

So our hypothesis was, well, and this is of course based on others work as well, not just this one study.

We thought, well, perhaps community onset bacteria caused by ESBL in the blood is actually occurring as outbreaks or are due to point source exposures. And if so, they should cluster by time and place. And if you look at the graph on the lower right, you can see this time trends, so the increase in prevalence of ESBL in the San Francisco public health care system where I am sitting now.

So our methods, so this is based on the same database that this prior study was done. So we obtained electronic health records from the San Francisco Health Network and General Hospital. Sorry, I'm giving a talk.

These are records from 2013 to 2020 and we obtained culture data. So anybody who had a urine culture sense, we obtained their data both for the urine culture speciation, drug resistance, as well as demographic characteristics and geo-coded addresses because this has been a major effort at UCSF to geo-code patients' addresses. So our sample here is focused on community onset bacteria. So this is represented by urine cultures that were obtained less than 48 hours within someone being hospitalized or in someone who was receiving outpatient care who did not have a hospitalization in the past 90 days. So this represents over 7,000 people.

A majority of them are women and in the age of 46 years and this is a multi-ethnic group.

So we mapped the distribution of bacteria caused by ESBL E. coli versus non-ESBL E. coli. This includes other drug resistant E. coli.

We also looked for clusters using global morons I and we tried to identify some hotspots using local morons I. I'm showing this to you but this is talking to a crowd that is very well versed in GIS and so you don't need those little formulas.

So in terms of distribution of bacteria episodes in San Francisco and I just want to point out that actual addresses are not shown.

We see that the episodes of ESBL E. coli and non-ESBL E. coli include so red being or orange being the ESBL E. coli kind of scatter around in the same distribution that we expect our patient population to distribute.

And when we looked at both, so the morons I index that I'm showing here is actually the global morons I index and we tried to identify hotspots. So here I want to be mindful to mention that we're looking at bacteria episodes and so there may be multiple individuals represented multiple times. So we see these clusters but this kind of didn't come at the right time. But when we decided to look at, well, what about if we just look at anybody who's ever had an ESBL E. coli causing a bacteria or UTI, you know, do those cluster and no they don't. Then we said, OK, so what about if we looked at smaller geographies? You may have noticed from, you know, this map that we do have areas of highly densely populated areas. So we restricted our analysis to the tenderloin and soma to the mission, Bayview-Hunderst Point and Excelsior. And when looking at the smaller geographies similar to what the Columbia paper did, we did not find clusters there either. And then when we decided, OK, well, if we remove patients that and maybe living at the same geocode or geocoded in addresses so that we do not have multiplication of either ESBL or just any UTI episode because they multiple episodes happen in the same person over this time period and removing the fact that multiple people may be living at the same address. And just looking at addresses, we did not find any special clusters either.

So what's going on here?

Our thought is that ESBL E. coli may actually be caused by multiple genotypes and there may be multiple point sources for this very general kind of drug resistance, even though it feels very-- it feels-- there's-- let's say that if someone is familiar with

microbit or nitrocarantin or a Bactrin, that resistance-- that prevalence of those resistance profiles is much greater. It's much greater than ESBL. But even when looking at ESBL producing E. coli, which is a very specific type of E. coli, we're not seeing that. So to add on to this analysis, so we collected urine samples from the general hospital, from Laguna Honda, and from San Francisco Health Network Clinics. We collected-- let me go quickly. We collected about 1,000 urine isolates from April to September 2019.

We connected those isolates to patient demographic characteristics and geocoded addresses. We then genotyped those E. coli. So about 551 samples were E. coli. And so we were able to look at specific genotypes and to see whether those actually also clustered. And this is a general map. This is not-- this is not the cluster analysis. And I'm not showing you more anti. But I do want to show you that in our latest analysis, ST95, which is represented in a maroon kind of brown color, does cluster. What's interesting about ST95 is that it is usually drug susceptible.

But we did find clusters. So our findings here suggest some local common source exposures for specific genotypes, like ST95, but not ESBL E. coli. And like I said before, this probably because ESBL E. coli is coming from multiple sources. So our future work is going to include identifying individual and neighborhood risk doctors or potential sources for individuals who have a UTI with ESBL E. coli or specific E. coli genotypes. And at this point, my team and I are discussing whether or not we should consider other methods to identify these clusters, looking at K function or even considering the newly request-home-based models that Hoonet offers. So I want to acknowledge my team at UCSF, my team of mentors, and my collaborators at UC Berkeley.

And I hope there's time for questions.

That was fantastic. Now we're getting a reverb there, but hopefully that'll be okay. So yes, we do have time for a few questions and also if any, because we have so many GIS experts online here, we have about six minutes for questions, comments, helpful advice.

If anyone has a question, let's see. I don't see any in the chat. I saw a hand raised.

Benji, you can feel free to unmute yourself to ask or add it to the chat and we'll read it.

Sorry to interrupt, Charlotte. No, that's fine because I'm not seeing the hand raised. So thank you, Myanna.

Who has the, there was a question?

It was. There was a question or there was a hand raised. It's gone now. I think maybe they raised an error or if they're on an iPad, an iPad, if you talk with your hands like I do, a little pop up comes up that says, are you trying to raise your hand? If you're on a Zoom.

Eva, that was so interesting. I guess, I mean, I've known Dr. Raphael for a long, long time. I knew her before she was doctor, so I should be in the habit of calling you Dr. Raphael instead of, okay, good. That was so interesting and I'm sure of great interest to the participants. I think your presentation was so clear that that's why there's no questions. It was very well spelled out in terms of both what the microbiology part of UTIs and also the analyses that you're supposed to do. I mean, one question I had and maybe a comment or some food for thought and potential future analyses of looking at time-space cubes and emerging hotspots might be something to scrutinize this data like that, of incorporating the fourth dimension into your cluster analysis could be interesting. And plus, it's just fun to make time-space cubes. So there's a couple of ways to do that, both with Esri products, which are always the easiest click and point and click, point and click, right? Or to program that out an hour, whatever suits you.

That's one thing. SAT scan could also be interesting. It will be really interesting to see whether you come up with the same sort of conclusions once you apply the different methodology, the SAT scan versus the Maran's eye and sort of the global clustering or the Getty's sword local clustering. So we'll look forward to that. And hopefully you'll get those publications to put out really fast and we can follow up on what you're reading.

I wonder if anyone else that's here, we have such a broad range of participants. If anyone else has some thoughts or advice of where Dr. Raphael might take her analysis from here, anyone want to contribute some thoughts on that?

Even an email with, here's what I think you should try because my team and I were like, oh, let's just try all these different ways to use Maran's eye. And it's hard to kind of say, okay, we have this, to have a strong publication if you're kind of talking

about negative results, but that is the nature of the data also. We were thinking of doing some temporal spatial cluster analyses, but Charlotte, thank you for pointing me to those analyses because there's so many.

And so, yeah, I would welcome any comments or ideas.

So compliments and clapping and the chat, but I think like I said, your presentation was so clear and well spelled out that we're not getting those questions. I often get, could you go over again? Could you say again or could you explain again? No, you don't need to do that. Everything was very, very clear. So go ahead. Sure. And the question was, if we looked at different backgrounds like race and ethnicity, because we do have patient race and ethnicity information. And so we were thinking maybe we might see clusters within ethnic enclaves as they do in ethnicity. And when we restricted to the specific one ethnic, we actually didn't see clusters either

in all these social categories.

So even by further defining our population, we didn't find clusters of PSPLE.

I'm wondering one thing because I've recently joined a gym that has a pool and all the ladies are talking about UTIs. Are you able to bring that question either using a common hot tub or a common pool, swimming pool into your analyses?

So we are planning some analyses where we will be obtaining further risk, like sub-patient recorded risk factors that you cannot get from electronic health records. We haven't thought to, you're right that I believe that pools have been questioned to be a risk factor in the past, but that might be a question.

You can actually look at that spatially too. Is there even a pool? Although people will drive to pools. But yeah, anyway, this was fantastic. Thank you so much. And we'll welcome our next speaker, which will be Alexandria Zermano at UC Berkeley.

Ethical Considerations in the Collection, Analysis and Reporting of Spatial Data

Speaker: Alexandria Zermeno, UC Berkeley

Abstract:

NA

Transcript: Video Timestamp: 00:19.33

Alexandria is an undergrad at Berkeley focusing her research on ethics in the geospatial sector. So please go ahead and share your screen, Allie.

Am I doing presenter view? You're doing presenter view. So if you go up in your scroll, your mouse up and you can switch that or.

Or even it's like a perfect. There you go. Perfect. I just showed all my slides. Hello, everyone. I'm Allie Zermano. That is my nickname. My full name is Alexandria Zermano. I am here to share about the how in geoethics, ethical considerations in the collection and analysis, collection analysis and reporting of spatial data.

And I'd also like to acknowledge that Charlotte Smith is also my mentor through this thesis. So I'm really excited about it. So to start off, I'm asking the question, what is spatial data?

And I mean, for us at UC GIS week, I'm sure everyone can pretty clearly answer this question, right? What is that spatial data is latitude and longitude. But to set the playing field pretty even, right?

Knowing that it's latitude and longitude, how many of us really understand and really know how ubiquitous spatial data is in our lives? Like who is gathering our data? What kinds of technologies are gathering our data? Who is seeing it? Who is reporting it, sharing it? How does this even affect policy and that is in our community, but even at a state or national level, global level even? And so we can think of spatial data as latitude and longitude. But once we start really asking these questions, it becomes bigger. And so then, which brings me and brings us to geoethics, which I thinking about the concept of geoethics can be very abstract and there's a lot of different ways to think about it. And this definition, I think really helped cleared it up that geoethics is ethics about how we ought to live together. The creation of an access to geospatial data means we live together differently than we did before.

So as we were saying, right, like our lives are very connected to spatial data. And that means also that there's a lot that we're connected to now. And even thinking about us as researchers or even us as individual citizens, community members, this is a big question that we really should continue to think about.

And thinking about even more so, right, knowing that spatial data is super ubiquitous in our lives, whether it be, and I repeat ubiquitous because I'm thinking of where when we're using our phones, we have weather apps, we're using our phones to get to different places on maps we are using.

There's BART systems, transit systems that are continuously tracking locations. And we are scanning our cards. There's a lot of different ways that our data, our spatial data is being tracked every day. And so thinking about the number and the amount of this data that is that we are as researchers or even people in a vacuum with a I found it really interesting, but the National Science Foundation funded a team from MIT. And they were talking and wanting to accelerate and grow the AI algorithms for processing data. So because there's these physicists and astronomers who are having such a large amounts of data, right?

And they even were able to see that this overcoming tsunami of data threatens to overwhelm huge data rich. And I wanted to like hide it a little bit, but it's research projects. The research with this overwhelming amount of data that is that we are interacting with, it's overwhelming even the research projects, us as researchers, and maybe for astronomy or for physicists, right? It's more about the computer, but even for us thinking about in the public health sense, thinking about the data that we're interacting with is more than just a leaf on a tree, but this is someone's personal information.

And so that's what even I wanted to share, right? Is that the, there is a, I don't know if this is blocking the view, sorry, there is a bottleneck. So I have like this, there was a bottleneck, right? Of trained and prepared researchers who are trained in understanding how to apply ethical considerations in human subject, geospatial data, because we are even able to see that

in the academic networks, network of Americas, the regional community committee of United Nations on global geospatial information management for the Americas, they asked, they created the survey even asking, are there any

standards on the ethical use of geospatial data? Because more than just us as researchers being able to, to like, in experience be able to apply these ethical considerations to data that contains human subject information.

And we, there's no, there, we were able to see that in the survey, 77% said no in their government agency, that there was no standards.

And so what many ended up doing, right, is thinking, okay, well, we need to build some principles. We need to build some foundations for these standards. And one of the, one of them is the Lucas charter.

But this leaves us with principles, this doesn't necessarily leave us with the how, right? We're here to think about, okay, how do we, how do we do this, even myself as an undergraduate, as a researcher, as a practitioner, I need to be able to apply these methods apply, apply addressing bias, applying, making sure that my, my participants have their privacy protected, that even, that even there is confidentiality, that they will be, feel comfortable with sharing their personal information with me.

And that's what leads me to this thesis, right, is, okay, how can we work on preparing and training ourselves?

So what we ended up creating is two surveys, a survey for institutional review boards, and a survey for geospatial practitioners.

And so what we're asking is, have you received formal training on how to execute ethical collection on geospatial data, for collection, for vulnerability, and for the institutional review boards, how, what standards or what guidance documents are you using to even make sure that we are being held accountable as researchers in reporting and collecting or analyzing spatial data?

And so that's where I want to leave us with is, even this survey, we're hoping to collect a shared breadth of knowledge of how practitioners are even applying ethical considerations into their work today. And so we are super close to getting our survey approved, surveys, both of them approved. But if it would be amazing if you guys could share your, if you guys are willing and interested to share your email.

And because so that way we can share the survey. Because what our ultimate goal is from even asking these questions, because I, I myself was starting, I started this project in the summer, working through STEER, an internship that Charlotte was my mentor in. And through the literature, I was able to find a good amount of methods that were super useful to share with others to share for myself, like, oh, there's this really cool thing called X, mask my XYZ, a tool David Swanland created to jitter points and help researchers be able to mask their information.

And but even this breadth of knowledge, right, I myself as an undergrad, and even I was finding that even many researchers still were meeting and really seeking out training as well. And especially knowing that this is very close to heart for all of us. And so if you guys could share your emails, if you guys have any questions and any methods, any ideas on on this topic, that would be amazing.

But yeah, like, I really appreciate to work with all of you and continue to work.

Fantastic, Kelly. Thank you so much.

So we I don't see any questions in the chat.

So, uh, my Anna has posted up the slack channel for first session to track to which we're in. And now you can feel free to if you're in the slack, that's the specific channel for this session. But maybe we can just open it up for a general discussion on ethical considerations of GIS and spatial analysis to help Ali with her undergraduate thesis project. I'm wondering if any of the attendees we have about 24 attendees here have any of you use those methods like jittering or masking or aggregation that you could share with Ali, some maybe use case or experience with that.

Feel free to take yourself off mute and jump right in. Is there any experience with aggregating data, perhaps single point addresses into a polygon?

Is there any experience with that among the group?

Hi. Hi, Anning. Go ahead, Liz. Uh huh. I think I may have a little bit like experience to working with that to having that experience, because I also work with survey data and with that survey that was like collected by individual reach out to individuals and they will provide their nearest home address. So they have a straight information and with that information, we first convert that to use a geo

coding, convert that to latitude and longitude. And then I also need to do analysis like on a census track level. And then I graduate, I aggregate them on census on to a census track. So I assign the latitude and longitude information to the census track that it belongs to and then make an aggregation for the people that within each census checked.

So that's all I do. Thank you for sharing that. And may I ask you, when you had very few actual geo coded addresses within a track, did you set a minimum number so that you would essentially define the polygon as zero if it was some minimum number? Or did you tell us about that? Yeah, I do have like threshold or limitation for that. So for my case, I put like below as five. So if the respondents, yeah, the respondents if in that census checked is below as below five, then I will put that as an A, something like that. So I will not include that census checked in my study.

So interesting. And have you published this work? Yeah, I'm going to share to present this work. Okay, you're the next speaker. So I don't want to don't do it now. Don't do it now. You're a 140. So thank you for the preview that I know everyone will stay on. And yeah, we're to hear about that. And that will be so useful to Allie. I know Allie will stay here for your talk with you know, at the edge of her chair.

Alexander, were you going to share something I saw you had taken yourself off of mute?

Sorry, she's having some issues with my zoom.

That's okay. No problem at all.

So that's fantastic. Yeah, this whole question, we all I mean, everyone that's probably come to UC GIS week is working with spatial data. And sometimes, yeah, like Allie said, if it's a leaf on a tree, we don't worry too much about ethical considerations. Although I know in Allie's work, she's also extending her work to sort of exploitation of the environment. So often people that are what we would think are more the natural resources as opposed to public health users of spatial data are also considering some of the ethical questions related to spatial data. So I'm excited and waiting for Allie's thesis next May to come out. She's at the beginning. Let's see if anyone was working in that field and knew of and wanted to respond to her survey. If you could put your email in the chat, then I'll pass those

on to Allie. If you're willing to, like I said, take her 10 minute survey about ethical uses of data gaining, I'm really hoping that you'll be willing to look at Allie's survey.

I have a question. Oh, let me ask a really quick, practical question. Allie, do you mind if I put your email address in the Slack channel too for people to give you their email and then I'll...

Okay.

All right, go ahead.

I have a question.

This is great. I didn't know some of these resources existed. I'm looking at housing displacement, risk and suicide mortality in San Francisco. And so I'm mapping, literally mapping those individual addresses. I wanted to ask, is there a central location to look for these geo ethical workarounds and they're much needed?

Go ahead, Allie. Yeah, there is. And I think I can... If you could drop your email, I can send some of the resources. Oh yeah, you did. Okay.

I can send some of the resources that I've been able to find, but yeah, I think that's what's really interesting about this project and even just the surveys, right? Because we're hoping to use that and once we get everyone's responses, we'll have the better-set answer. Because there's some of what everybody's using, right? Rather than maybe just one person, but this is the most useful method.

Right. Because my challenge is I'm looking at only a couple of years of data and suicide is not super, super, super, super common. And so it's hard to set a threshold because you don't really want to lose any data points for that. So that's what I'm struggling with in terms of aggregating. And so I just... Yeah, so I would love to see what the options are.

Yeah. Yeah.

We can share that and I see you put your name in the chat. I mean, UC GIS is all about networking, sharing, sharing resources, learning from each other. So I'm excited that we can share some things with you, Kinshata. And these days I'm actually a student at Johns Hopkins doing the Masters of Applied Science. And actually this week in class, it was all about suicide, suicide prevention. So I'm going

to send some links to articles that we had to read as homework. I'm just a lowly student again too, but your question is timely because that was this week's lesson. So be looking for some things from Allie. She'll share the locus charter. So you can see sort of one of the tenets and principles of the locus charter. And then she's been finding...

So the locus charter is what Allie's really focusing on, how. How do we protect human subjects? Yeah. Like your project too, one of the examples that I ran across was like they did... It was Hurricane Katrina. And so they specifically mapped out, "Oh, here's the mortality of like from this hurricane." And then when they reported this map, the locations of the people were on the map still. So like...

So even from just taking that piece of paper, we're able to reverse, do code it or even find those people, find those homes. We went to... When this person died, when they came and picked them up. So even just thinking about that, right? And especially about suicide.

That's very serious. And that is a very big...

Great. So Dennis, I see you've shared in the chat as well. So we'll put you into our small network here. And Allie's survey is going to go out through some of my networks at United Nations, GGIAM, and Organization of American States Page Group and Friends at Census. So we're trying to disseminate this widely and those that want to stay sort of in touch with Allie. She put her email there as she starts to write her thesis. Hopefully we'll get that published as well about really applying these ethical considerations to our GIS data collection as well as spatial analysis. So stay tuned.

And I see Menchie, you're also sharing your emails for Allie as well.

Evaluating health outcome metrics and their connections to air pollution and socioeconomic vulnerability in the Coachella Valley

Speaker: Yaning Miao, UC Riverside

Abstract:

We use two different measures of health outcomes at the census tract level – one from randomly sampled telephone surveys and the other from emergency room visitations – to investigate the degree to which health outcomes may correlate with criteria air pollutants including fine particulate matter (PM2.5) and ozone, and socioeconomic metrics such as education level, linguistic isolation, poverty, health insurance coverage, and racial demographics in the Coachella Valley. We further investigate biases and differences between different health outcome metrics themselves.

Transcript: Video Timestamp: 00:39.30

Thank you so much. So we'll move on to our last speaker of the session, which is Yanying Miao, a fourth year PhD candidate at William Porter's group in the Department of Environmental Sciences at UC Riverside. Her research is mainly focused on air pollution modeling, health effects, and environmental justice.

Go ahead, Yanying, please share your screen. Thank you. Thank you very much.

So good afternoon. I'm so happy to be here at the UCJS week. And I'm Yanying Miao. I'm a graduate student working with Dr. William Porter from UC Riverside. So today I'm going to share our research work that has been done in the Coachella Valley and to evaluate the health outcome metrics and their connections to air pollution exposure and also the socioeconomic vulnerabilities.

As we all know that the unmet air pollution exposure is one of the leading threats to human health. And beyond their overall, beyond their general adverse human health effects, their impacts also shows disparities among different socioeconomic levels. So, for example, on this map, we are able to see that areas have a higher proportion of black people, tends to have relatively high PM 2.5 achievable mortality rates. So in general, people with lower socioeconomic status, they tend to experience with higher air pollution levels and also associated health outcomes. And these people, they live close to some major air pollution sources. And one of them is the inland lakes, or you could say the shrink of inland lakes. And Southern Sea is one of such inland lakes that has been shrinking for many years. And it already caused the air quality and also public health concerns for its local residents.

So the Coachella Valley is an area that contains several cities located at the north end of the Southern Sea. And it has experienced very high levels of both air pollution and also just those economic vulnerabilities by both the national and state level standards. So, for example, Riverside, sorry.

So, for example, Riverside-Kangi is regularly designated for serious non-attainment area for the coarse particle, which is PM 10. And this is in large part due to the dusty conditions in the Coachella Valley and also other desert areas. Beyond these facts, the disparities in the health outcomes and the socioeconomic status are also very apparent within the Coachella Valley itself. So, for example, the lowest asthma emergency visitation rate is around 14 per 10,000, while the highest will be reached as around 72 per 10,000. Also, the poverty level range is very broad and start from around 70 percent to up to 87 percent. So with this with this fact, we hope to examine the health outcomes and their connections to the chronic ambient air pollution exposure and also the socioeconomic metrics. We also want to investigate the biases and differences between the different health outcome metrics themselves. We have policies that the spatial patterns of the ambient air pollution exposure levels will be positively related with the health outcome metrics.

Also, the socioeconomic vulnerability will exacerbate the health consequences of this exposure and further contribute to some negative health outcomes.

So in this study, we are including three different types of data. The first day is air pollution data and our target air pollutants are PM 2.5, which which also known as a fun particulate matter and ozone. We also have our health data and we're including two different health metrics in our study. One is survey data, which is collected by local organization located within the Coachella Valley called HARC. And the other one was the emergency visitation data, which was collected by the state public health department. And at last we have our socioeconomic status data for the socioeconomic status data that on the census tract level, we use the five year American community survey data. And for the socioeconomic data at the individual level, we use the indicators directly from the HARC survey data. So with this different data size, we're able to conduct some statistical analysis and also data visualization to examine the spatial and also demographic specific patterns of the air pollution exposure and also the health outcomes. So here are some major results that we have right now. So these maps showing the spatial distribution

patterns of the five socioeconomic indicators on census tract level. So on this map, we are able to say that we found that there is a very clear spatial patterns for the socioeconomic status. So typically, the high socioeconomic vulnerable communities are more likely to cluster around the east and also the west ideas of the valley. So it's very similar for all of them, five. So it do have some disparities for the socioeconomic vulnerability.

So here on this map, you see are the spatial distribution of the PM2.5 and ozone. So we can say that it seems like these two air pollutants, they are coming from both sides of the valley with very strong patterns from the east side with a very high PM2.5 levels from the eastern side and with a very strong ozone levels from the west side. So this is this is very clearly formed a very strong spatial contrast of the ambient pollution exposure. We also separate our air pollutants by the socioeconomic status level, as you can see from the box plot on the right side. So in general, we found that communities or census tracts with higher socioeconomic vulnerability, they are more likely to experience with a higher PM2.5 levels, but lower ozone levels. So it's kind of indicating another disparities for the ambient air pollution exposure.

And then we examined the health metrics and we first examined the emergency visitation data, or we can call it e-mail rates. So here on this map, you see are the spatial distribution for the e-mail rates of asthma and cardiovascular disease. So on this map, we also see a very clear spatial patterns, which which which indicating that the higher e-mail rates are you already found also on the edges of the east and west edges of the valley, which. So this matches matches up with our with the previous findings that we have seen for the air pollution and also for the socioeconomic data data. So with this emergency visitation data, we did linear regression analysis, used e-mail rates as the dependent variable, used air pollution, socioeconomic status and also several interaction terms.

According to the results, the e-mail rates shows a very strong and significant positive relationship with air pollution, with socioeconomic status and also the interaction term between air pollution and socioeconomic status. And we also see a very strong and significant negative relationship between the e-mail rates and the interaction term of PM2.5 and ozone. So this indicating that in the areas where you have a much higher ozone pollution, the impact of PM2.5 on the e-mail rates

will be reduced. So I think this makes sense to our study area. So if you recall the spatial distribution maps of the air pollution, you will remember that the areas, the areas where it has a very high ozone pollution, the particle levels are relatively low in that area. So this makes sense that the impacts of the particles are not so strong. So in general, the results from the emergency visitation data meet our expectation. However, the e-mail data may not tell the full story because it's like a very extreme situation. I mean, the emergency visitation data, it doesn't include the people who suffering some symptoms, but they don't need to go to the emergency department or they don't have access to the emergency service. So in this case, we introduced our survey data, which includes many different types of medical behaviors, not only limited to the emergency department. So here are the spatial distribution of the survey rates for asthma and a heart attack. And in generally speaking, the spatial, we don't find a very clear or strong spatial pattern for the survey rates. It's kind of relatively more spread throughout the whole rally compared to the e-mail rates.

Since our survey data was reached out to individuals and we totally have around 2000 respondents for the survey. So we decided to use logistic regression analysis to do a logistic regression analysis, to do analysis on the individual levels for the whole respondent. We also use a weighting factor to avoid some biases that caused by some factors like age or sex, something like that. And according to the result, we found that the survey rate has a positive relationship with PM2.5 and ozone, which is similar with what we have found for the emergency data. However, when we look at the correlations between the survey data and the socioeconomic status, we found a very interesting situation. According to the result, the respondents who identified them as Hispanic, who have linguistic isolation, they tend to show a very lower rate of the survey reported health outcome. And also the health inference stands out as consistently strong positive indicators for asthma and heart attack. We also found a very similar pattern for the whole respiratory disease and cardiovascular disease. So this is kind of an unexpected situation for us because we always expected that higher vulnerable people are more likely to have a higher disease rate. So we can also we start to explain, explore the survey data and try to explain this situation. And there are, we found several, a couple of reasons may explain for this situation. The first one is the survey question that they used. They used in that survey. So the question they asked to collect the health

clinicians is, have you ever been taught by a doctor or medical professionals that you have the following disease? So this means that only the people that who has visited a doctor and diagnosed the with and diagnosed the bioductor would be treated as they have such a kind of disease. So this in this case, you will miss the people who also suffering from some symptoms by the neurovisit of the doctor. So this is probably one reason. And then the other one is that maybe caused by the sampling biases. So we compared the socioeconomic status levels between the survey respondents and also the surrounding census check. Also, their surrounding census checks. And we also examine the respondent distributions for at the socioeconomic status levels for the within the respondents. And in general, we see that the survey response, most of the survey response, they came from a very low socioeconomic vulnerabilities and they kind of, they cannot represent. They have a very clear and they have a very clear biases to the lower socioeconomic vulnerability of people for their response for the respondents. So this means that we may have some sampling biases for the survey. So that's all for for our study. So some major conclusions. So the first one is that we say a very strong patterns and disparities in our pollution score, health outcome metrics and also the socioeconomic vulnerabilities. And both we say both PM2.5 and Ozone shows positive relationship with both the emergency visitation data and also the survey reported health outcomes. And for the relationship with socioeconomic status, we found that the emergency visitation data shows a positive correlation with socioeconomic vulnerability. But the survey reported health outcomes indicated a negative association with the socioeconomic vulnerability. And we think

the sampling biases in the health outcome this side, the survey question design and also some other factors like visually significant influences may contribute to the biases that we found in the survey data. So in our future, our future work will be would like to examine these differences and apply them to the analysis of some acute does the UN exporter as well as the method for some risk communication for the Coachella Valley residents. So with this, I would like to send to my PI Dr. William Porter and also our collaborators, Dr. Gershwabi and Dr. Jenna Likot Henik. So and also the founding from NIH. Thank you.

Thank you so much. This was fantastic. I have to tell you, there's a graduate student at UC Berkeley who wasn't able to attend, but thankfully these are

recorded and she's going to be, you'll be her hero because this is very much what she's interested in.

I see you're getting some comments in the chat thanking you for your excellent presentation. Look here at the chat to see if there are questions.

Still some thoughts on Ali.

I don't see any questions in the chat. Oh, there's one. Is there a way to obtain smoking data?

I think in the survey, we do have like a question asked whether they have been a smoker or not. And this factor was adding to our weighting factor to weight for the whole population.

Excellent. What's next in this research? Shannon Langley is asking.

Okay, so the next steps will be that we will like, so in this research, we combined two different health metrics. One is hospital data and one is a survey data. So by in the current research, we just like

an sized house correlations for the whole like the air pollutants. But in our next, we are going to, so another situation is that in the South Sea area is pretty dusty area and the PM 10 is a very is one of the major air pollutants. So in the next work, and I took folks on the dust events, and then we try a way trying to use trajectory simulation trajectory model that is called high speed and combined with some work simulation weather data to spy, see why or identify the South regions for different dust events. So we will separate the dust events from different regions, from different directions because different directions will cause the dusting events have different chemical compounds and will cause like a little bit different. We hope to see a little bit different health effects of the dust to coming from different South regions. So this is our next step, we would like to identify the health effects of the dust from different South regions in the also still in the Coachella Valley.

Fantastic. We're at about the end of this session, but if you could stop share and then my Anna will put up the screen that will show the survey q qr code if you can if the attendees would be willing to give us some feedback using the can you do stop share.

Oh, yeah.

Then if the attendees could give us feedback using the QR code. And remember we have the slack channel and our UC GIS website to that's a way to stay in touch with all of us, the new people that are your meeting here and please consider presenting or being on the planning committee for next year. I don't know what COVID will bring whether we'll be back on our own campuses or do another 10 campus wide UC GIS. So thank you so much for your attendance at two o'clock which you can now go to the other room for public health to if you're interested. Thank you so much, everyone.

There are some chats throughout with lots of interest in data and geo ethics discussions and follow up so we can also continue the conversation in slack to if you're able to join. Exactly.