So, so first it is my honor to today to welcome Doctor. She only as our summer speaker, Xihong Lin is a professor from Harvard, jointly appointed by both as deaths. Annabelle Slash Department and she has received broad recognition and many awards for her great contribution to the field and her research has covered so many different topics ranging from Mr Logical work including hypothesis testing, had dimension statistics and color inference tools, data applications.
And the computational statistics such as Statical Genetics. Scalable statical inference as well as applications with epidemiological and health data. So today she will share with us her work on analyzing large scale coordinating databases from both China and the US and provide several takeaways and discuss priorities. It’s Insp endemic, so I will not occupy everyones time for more. I will hand it over to see home from here. So should we start? Thank you laying so
much and for inviting me and for the very nice introduction and so I will share my screen.

1st oh, I cannot share the screen. I have just met your Co host so you should be able to get it now. Can you can you share your screen I can do now? Thank you thanks. Alright, can you see my screen? Yes. Cool excellent alright?

So I’ll share with you some of the work and we have been doing last year on the Covid 19 so this is the data just downloaded earlier this week. So as you can see that the right now
they're over 110,000,000 cases in US and also 2.4 million deaths worldwide.

So if you look at the curve on the left, that is the case curve and for a few selective countries. So as you can see for both the UK and also United States and the number of cases had been going down in January, so that is a good sign.

And also Israel as you know that Israel has been a really leader in a vaccination and so their cases have been going down as well.
But if you look at Africa I think the number this particular country you can see the number of cases. Has been going up likely due to the new violence in Africa and on the right. That is the case curve with the desk curve. You can see the patterns pretty similar, especially you can see for this African country and the number of deaths has going up quickly, so that is really worrisome.

So here’s a talk outline, so I’ll start talking about the covid transmission intervention and using the data, and then they turn talk about the USN word data.
Then I’ll talk about epidemiological characteristics of Mackovic.

Then I’ll talk about the 221 playbooks and also the defining challenges in particular about the vaccine rollout and uptake. Our focus more on the uptake and also how can we do scalable testing? And in particular, I talk about this support design and we call the hypergraph we called hyper. That is based on the hypergraph factorization.

So I started working on the Covic research mainly by coincidence.
and last February.

So my post, our former postdoc column one

he is currently is professor in school public Health at Wild on Science and Technology University, which is located in Wuhan, and so last February,

’cause Wuhan was epicenter. So I wrote him a message asking how and he and his family were doing and he told me that. He and his colleagues were analyzing the Wuhan data, then at that time there was already one case, and in Seattle and one case in Boston.
So I sense that the virus might spread and so.
I decided to join them and working on analyzing the Wuhan data so we work for several weeks and in February especially child when his Cody worked really hard and then finish this preprint and we post the preprint on March 6 and so then with the hope that the findings we want we want to share their findings with the US and also other country as soon as possible to help them.
preprint attract lots of attention

that you can see the number of abstract view and download,

and also there lots of. A lot of free trade and also the news outlet coverage and also the policy documentation.

And after I twittered this paper in the Twitter and so then this paper preprint cover too much stuff and so we decided to split the paper into split the preprint into two paper and one was published in JAMA and last April. Another one was published in Nature last summer and so this work on the JAMA paper was led by. I’m Pan and a towel,
and here they both were addressed.

pH alarm and country.

Who is the Dean of school?

Public health as well doing science

University and the

Nature paper was led by Charlo.

And then emerge after the preprint

was posted them in Twitter.

And then I got quite a few interview request.

But as a station,

my first reaction was turned

down all the interviews,

so I turned on all the interviews

and in March, and then April,

I decided that probably is not a good idea.
It’s good to talk with the media and then they can understand the scientific funding correctly. Then that will help, and the combat the Covic. And so I decided to accept the interviews and you can see there’s a quite few coverage of the findings and also in the US medias. And also in the UK medium and also in scientific journals such as Nature and Science. And the one thing I learned was that ’cause when as a faculty member in Academic Institute and we were not trained to.
00:07:00.780 --> 00:07:03.868 Speak to the media and so to the
00:07:03.868 --> 00:07:05.458 scientific communication on the two.
00:07:05.460 --> 00:07:07.620 The general public is very important,
00:07:07.620 --> 00:07:09.744 so it’s important to have more
00:07:09.744 --> 00:07:11.580 training and in this area.
00:07:11.580 --> 00:07:13.770 Another thing I learned was is
00:07:13.770 --> 00:07:15.712 important that speak in simple
00:07:15.712 --> 00:07:18.148 language and without the jargon and
00:07:18.148 --> 00:07:20.590 so general public could understand.
00:07:20.590 --> 00:07:23.446 And then I testified the in the
00:07:23.446 --> 00:07:25.247 science and Technology University
00:07:25.247 --> 00:07:27.243 $0.10 at Technology Committee
00:07:27.243 --> 00:07:30.040 of UK Parliament on April 17.
00:07:30.040 --> 00:07:32.668 And so this sense technology Committee
00:07:32.668 --> 00:07:35.411 has about 8 to 10 Parliament
members like a senators on it.

And then they later on they also invited few other witnesses.

And then they wrote a letter to the Prime Minister Johnson and they make 10 recommendations in their letter.

Under so then I was honored that several of my recommendation we included in their recommendation in the recommendation they made it to the Prime Prime Minister Johnson.

And then, uh, this last month and as a one year anniversary of the colic outbreak and nature published their third editorial focusing on the major findings.
a key finding from ideology. And so we will. We will honor that our paper was featured on the in this editorial. So let me first talk about the Wuhan analysis. Then we talk about the US and other countries and the 1st I'll introduce this effective reproductive number. So this concept, our value is right now and everybody understand what that is and so that measures average number of infected people by one case. So you can see on the right if
Artie called before that means one person could affect to four people. That means the virus spread. So therefore, to in order to control the pandemic, the army to be less than one. And so here’s a woman analysis under. So we found the two major features of Kovik. The first feature is the virus is highly transmissible, and so the before January 23rd, and there was no intervention. As you can see, the number of cases had went up quickly, exponentially, and so the first case was reported.
00:09:34.012 --> 00:09:37.244 on the in this orphan and seafood
00:09:37.244 --> 00:09:39.504 market and this seafood market
00:09:39.587 --> 00:09:41.647 was closed on January 1st.
00:09:41.650 --> 00:09:44.786 Under then on March on January 23rd,
00:09:44.790 --> 00:09:50.370 and that was a lunch of the lock down and so.
00:09:50.370 --> 00:09:56.799 As you can see that before the lock down
00:09:53.592 --> 00:10:00.104 so that means each purse case could be.
00:10:00.110 --> 00:10:03.809 Like 3.5 people and so that is not good.
00:10:06.393 --> 00:10:10.970 and then after the lock down and
00:10:06.393 --> 00:10:08.009 disease is very transmissible
00:10:08.009 --> 00:10:10.970 and then after the lock down and
00:10:10.970 --> 00:10:13.171 with the social distancing is
00:10:13.171 --> 00:10:16.132 really how the RT value dropped to
00:10:16.140 --> 00:10:18.606 a little over 1.2 and then.
But it’s not good enough and so then after February 2nd and that was a lunch of the centralized isolation and quarantine.

So they basically the city build. Two new hospital field Hospital and also 16 field hospitals converted from the stadium and expectation Center and after that.

And then you can see the number of cases drop down very quickly and the RT values and by March 8th was about .27 and then the pandemic was controlled.

And so to estimate those RT values. So basically the model we use is a person partial differential equation model and so this captured by the left here.
Then you can see here is a from them. So here this is the symptom onset and between exposure and symptoms answer discord, incubation period. So generally this is about 5 days and then from the exposure to presymptomatic onset. This so this part period called latent period. So that means a patient is infected and but the person is not transparent, doesn’t transmit the disease, and so. Between the pre symptomatic period and the person, even though a person doesn’t have.
a symptom and but very load, 
NOTE Confidence: 0.8317293

is high enough and then the person 
NOTE Confidence: 0.8317293

could become. 
NOTE Confidence: 0.8317293

Infectious and so this period is 
NOTE Confidence: 0.8317293

about two days between presymptomatic 
NOTE Confidence: 0.8317293

two symptomatic and between 
NOTE Confidence: 0.8317293

exposure to pre symptomatic period. 
NOTE Confidence: 0.8317293

This is about 3 days, 
NOTE Confidence: 0.8317293

and so we built in those components 
NOTE Confidence: 0.8317293

and in the model, 
NOTE Confidence: 0.8317293

so we introduce this 
NOTE Confidence: 0.8317293

presymptomatic compartment, 
NOTE Confidence: 0.8317293

and also because at that time the testing 
NOTE Confidence: 0.8317293

case will not that widely available, 
NOTE Confidence: 0.8317293

there were a lot 
NOTE Confidence: 0.8317293

of cases which were uncertain,
and therefore we built in this. Unstained components, the observed data here those are the observed data and then those are. Then we construct all those components and drew the deep partial differential equation. So this is isolation component. And so so here the we after you feed this model, one can construct the reproductive number or value. So basically, as I mentioned, the datas are here and those are the data. The rest part are basically coming.
from the partial differential equations and then we fit the partial differential equation, Apostle model and using MCMC and the Gulf Coast parameter estimate. And so here you can see that before the intervention on our journey, then if there was suppose there were no intervention, then this will the blue curve will be the predicted number of infected cases and so you can see that 170% of the Wuhan population were infected. That will reach the natural herd immunity and so now has a 10 million population size.
Then that means 7,000,000 people need to be infected in order to reach the herd immunity and so this is not a good strategy and many old people will die and the number of cases still going up but not as fast as the without intervention but after the. Centralized isolation quality on top of the social distancing and the number of cases went down very quickly. And so this is what they
did in one hand.

So if the case subject was tested positive and this person to patient was admitted to the field hospital, and so this is different from the US in the sense that in the last spring and US also build multiple field hospital, but they only admit as a severe diseases, so amount and diseases cases they were isolated at home. And so that these notes are isolation at home could still infect the family members. And but in Wuhan all the mild cases and were admitted to the field hospital, so they were monitored.
If anybody became severe and the patient was transferred to the ICU, the regular hospital and for the people who had a symptom and but who? Because at that time there were not enough testing kits and so they were quitting the odds of hotels and will University dorm. If so, all the children and stay with the parents so the family were together. If anybody became a test positive and the person was transferred to the field hospital in two weeks, if a person was tested negative and the person went home,
and similarly for close contact and they were quarantined as a hotel as well. I see how can I ask a very basic question. If I look at the Group One Group, Group One is confirmed. I guess if it means if you perform some kind of PCR test you are positive and Group 2 is with symptom but not confirmed. Can it go the other way? Can you first be confirmed but with no symptoms? Oh yeah, I think there were cases. Possible that would have had no symptoms but test positive, but at that time because they were
not enough testing kids so many cases and who were able to be tested at the same time, so that’s why there were lots of undetected cases. I said thanks.

And so this strategy worked quite well in Wuhan. So in less than two months they reach 0 confirmed case and then by March 18 and by April 8 and after. To confirm the case for three weeks and then the city was reopened.

So it’s a whole thing only took in two months and so the first
take home messages,
NOTE Confidence: 0.8223038

the social distancing and centralized
NOTE Confidence: 0.8223038

isolation quarantine were critical
NOTE Confidence: 0.8223038

for controlling the outbreak,
NOTE Confidence: 0.8223038

so using the social distancing
NOTE Confidence: 0.8223038

alone that help but was not good
NOTE Confidence: 0.8223038

enough so that helped make our
NOTE Confidence: 0.8223038

reduce around 1:00 and but did
NOTE Confidence: 0.8223038

not bend the curve and the reason
NOTE Confidence: 0.8223038

is there were lots of the.
NOTE Confidence: 0.8223038

Community transmissions on the
NOTE Confidence: 0.8223038

social distancing help block the
NOTE Confidence: 0.8223038

community transmission that is
NOTE Confidence: 0.8223038

between household transmission but
NOTE Confidence: 0.8223038

within family transmission and cost.
NOTE Confidence: 0.8223038

Place transmission with common
and so the social distance.

Distancing does not help block that and so, especially with how many families are multi generation families and they live in apartment so it compared to US as even harder under to isolate. At home.

And the idea that centralized isolation, creating to social distancing that help bend the curve and stop the pandemic. And so we validate those findings and in other countries last spring.

So if you look at the curve on the left, that is a Italy data.
the social distancing that help reduce
NOTE Confidence: 0.8223038
the R and the R curve lingered around
NOTE Confidence: 0.8223038
one for over a month and did not bend
NOTE Confidence: 0.8223038
the curve but and also on the right
NOTE Confidence: 0.8223038
that Germany data in the spring.
NOTE Confidence: 0.8223038
The same thing under the curve did not band.
NOTE Confidence: 0.8223038
And the second feature from
NOTE Confidence: 0.8223038
analyzing the Wuhan data is the
NOTE Confidence: 0.8223038
Covic is highly converged on the,
NOTE Confidence: 0.8223038
so we estimated about 87% of
NOTE Confidence: 0.8223038
the cases were undetected.
NOTE Confidence: 0.8223038
So in other words,
NOTE Confidence: 0.8223038
the detected cases was only
NOTE Confidence: 0.8223038
the tip of the iceberg,
NOTE Confidence: 0.8223038
and so you can see that on the on
NOTE Confidence: 0.8223038
the left and the right bars are
the detected cases and then the.

Yellow bars also uncertain the cases, so we estimated on the we estimate

about 87% of the cases were uncertain, and many of those cases were asymptomatic

By adding the yellow and red that can give us a prevalence estimate that is about 2.5% in one hand, and so this is similar to the theological studies based on antibodies, and that was about 3%.

And then US result very similar.

the CDC did theological study last
and then the estimated about 862 twenty times the number of cases were six to twenty time of the cases which were reported. And also those undetected tasted post a high risk of resurgence if one reopened too early, lifting the controls, and so we estimate the probability of the researchers. Think about this is the first day and one has a confirmed case. It doesn’t mean there is no case at all because there are still a lot of undetected cases. And suppose when we open in 14
00:19:46.636 --> 00:19:48.687 days by lifting all the control
00:19:48.687 --> 00:19:50.865 measures and the first strategy is
00:19:50.865 --> 00:19:53.109 after the first day observing the.
00:19:53.110 --> 00:19:55.200 Zero confirmed Case No matter
00:19:55.200 --> 00:19:58.034 whether the second day has the case
00:19:58.034 --> 00:20:00.358 or not and when to reopen info.
00:20:00.360 --> 00:20:03.896 This and the second strategy is one has
00:20:03.896 --> 00:20:07.559 a confirmed 0 case for 14 consecutive days.
00:20:07.560 --> 00:20:10.520 That basically means 000 our
00:20:10.520 --> 00:20:13.480 way through and what is.
00:20:13.480 --> 00:20:14.890 Research is probability.
00:20:14.890 --> 00:20:19.060 So that is what we found that if one
00:20:19.060 --> 00:20:22.196 reopen in 14 days after the first day,
00:20:22.200 --> 00:20:23.175 observe 0 case.
00:20:23.175 --> 00:20:26.238 So that means it can be zero and 120
again in this type of situation then the researchers probability is 97% and if one observes the zero case for 14 consecutive days and then the resurgence probability is 32%. So what is tell us is we need to be management and don’t reopen too early. Then we saw those cases are searched in the summer in the South. So what’s the take home away? Take a take away message on the number 2 is to control the pandemic. A single control measures not enough
NOTE Confidence: 0.84831905
00:21:07.970 --> 00:21:10.919 money to use multiple control measures,
NOTE Confidence: 0.84831905
00:21:10.920 --> 00:21:12.732 and including the mask,
NOTE Confidence: 0.84831905
00:21:12.732 --> 00:21:14.544 wearing social distancing and
NOTE Confidence: 0.84831905
00:21:14.544 --> 00:21:15.450 massive testing,
NOTE Confidence: 0.84831905
00:21:15.450 --> 00:21:17.700 contact tracing and also supported
NOTE Confidence: 0.84831905
00:21:17.700 --> 00:21:19.950 isolation and quarantines and also
NOTE Confidence: 0.84831905
00:21:20.016 --> 00:21:22.686 effective treatment and also the vaccine.
NOTE Confidence: 0.84831905
00:21:22.690 --> 00:21:25.525 And so the in the JAMA paper
NOTE Confidence: 0.84831905
00:21:25.525 --> 00:21:28.309 we call it multi faceted.
NOTE Confidence: 0.84831905
00:21:28.310 --> 00:21:30.734 Intervention and then later on in
NOTE Confidence: 0.84831905
00:21:30.734 --> 00:21:33.809 the summer and people give it a nice
NOTE Confidence: 0.84831905
00:21:33.809 --> 00:21:36.400 name and called the Swiss cheese model.
NOTE Confidence: 0.84831905
00:21:36.400 --> 00:21:39.586 So that is a nice name and so the
NOTE Confidence: 0.84831905
00:21:39.586 --> 00:21:42.090 challenge is we we know those
NOTE Confidence: 0.84831905
00:21:42.090 --> 00:21:43.326 in control measures,
but it’s difficult to implement

those control measures and also keep

high compliance in many countries.

So the defining challenges the

measure implementation and then

keep up with the compliance and

also the vaccine definitely is.

Really critical and we.

We know that by now there are

two successful ovac seen,

one in US1 is the face by Pfizer,

the other is more donor under

the efficacy is 95%.

This really really amazing

scientific advance.
Developing the vaccine in such a short time.

Last spring we also developed a website that helps estimate the RT value as a different resolution at the for different countries, States and counties and so you can see that. This work was led by Andy Xu, my student, and so this website was featured on the in Nature Article last summer. And so this website was featured on the in Nature Article last summer. So how do we fit this model?
So because there are lots of data points, we want to estimate the curve so therefore instead of using the partial differential equation model and we extended this epidemic model which was originally proposed by query in 2013, and so the model in this type of epidemic model is quite different from the traditional logistics model. So we need to build in the infectious component, If the number of cases is so think about the number of cases and for each day in Connecticut and then why is the number of cases so think about the number of cases and for each day in Connecticut and then
00:23:42.018 --> 00:23:44.374 one first need to calculate this

00:23:44.374 --> 00:23:47.286 Lambda T and this Lambda T is called.

00:23:47.290 --> 00:23:48.742 Basically calculates the number

00:23:48.742 --> 00:23:50.194 of people at risk,

00:23:50.200 --> 00:23:52.342 so that is calculated using the products

00:23:52.342 --> 00:23:54.569 of the serial interval distribution.

00:23:54.570 --> 00:23:57.412 Multiply the number of cases in the

00:23:57.412 --> 00:24:02.690 previous period said like 7 days and

00:23:59.946 --> 00:24:02.690 then the Ark is a parameter one moment.

00:24:02.690 --> 00:24:04.793 Estimate so in the original model,

00:24:04.790 --> 00:24:07.093 the estimate RT at each time point

00:24:07.093 --> 00:24:09.338 that estimate a lot of parameters,

00:24:09.340 --> 00:24:11.275 and then when building this

00:24:11.275 --> 00:24:13.874 person model and with Lambda T as

00:24:13.874 --> 00:24:15.638 offset and RT as a parameter.
But they asked me lots of parameters and one also account for the reporting delay by using a lag. So we did here with we try to accommodate on the Covid features and so we estimate us zero interval distribution and from this comma distribution using the paper in publishing in nature method and then in order to estimate RT as many values and we assume a curve and the estimate by using a spline. So there are few angle in work and so we want to estimate RT as a function but cover it and also the traditional epidemic model.
One assumed answer. Him and trade is a constant. Overtime and so the, but in practice the entertainment rate is not constant, especially when the number of tests goes up. What number of positive test rate goes up and then uncertain manner it will get better and.

So we want to answer payment way to be a function of the coverage and also we want to instead of fixing the reporting deal if we want to use the data to model the reporting delay.
and using all those met component, we can estimate the prevalence. So here are some preliminary result and so.
The code so you this is for the US data you can see right now and many countries a number of cases between being going down really nicely and so the current USRT value is about .78, and we hear when you can see we have this arty curve that expanded so below 1 now. And also you can see the number of new cases have been going down and also the number of deaths has been going down. But there is a lag between the best.
00:26:13.324 --> 00:26:15.310 the state level are key value.
00:26:15.310 --> 00:26:17.718 So just give example like for California
00:26:17.718 --> 00:26:20.643 you can see that the art in California is
00:26:20.643 --> 00:26:23.585 about .67 and so does this very nice banded
00:26:23.585 --> 00:26:28.018 curve for the cases and also for that.
00:26:28.020 --> 00:26:31.070 So now let me talk about the what are the
00:26:31.150 --> 00:26:34.130 factors associated with Covic infection.
00:26:34.130 --> 00:26:37.378 So as we start from the Wuhan data,
00:26:37.380 --> 00:26:40.229 then I’ll move to the US data.
00:26:40.230 --> 00:26:42.732 So the data we estimated the
00:26:42.732 --> 00:26:45.109 attack rate on the my age.
00:26:45.110 --> 00:26:47.546 So you can see that the each
00:26:47.546 --> 00:26:49.590 of the period separately,
00:26:49.590 --> 00:26:52.534 and so you can see that for the
00:26:52.534 --> 00:26:55.290 older people that purple and yellow,
and then the tax rate was.

Much higher than the younger people,

And in the spring last spring,

then later on, as you know,

like in US and there were more cases.

Elderly cases in the spring and but

then the elderly is become very careful

and try to protect themselves and

most of the cases in the summer and

also in the fall were younger people.

And then on the right that shows that

the male and female that you from one

day to the attack rate was similar.

But health care worker,

the purple bar has much higher
infection rate, especially before the intervention, and then after interventions. Acrid among the health care worker, and was better and so that calls for the importance of the PPS and before the intervention. People were not aware of the Covic, and so, therefore, is this not many people, not many health care workers. Hard to pee pee. So I give a talk on the Wuhan finding on the March just before the school public. Just before Harvard started the spring break and throw in one of the slides,
I showed that the on the day before the ABC News, there’s one picture of the health care workers. And so I showed up there so the health care worker will not properly protected in US and so they had no protection suit and no face shell for example. And then the infection could be go through eyes. I did not realize that those three slides on the showing the health care workers not properly protecting US were widely distributed during the weekend. So the March 13 was a Friday.
Then on March 16, that was a Monday, and so there was a national campaign on the protection of health care worker, which comprehensive PP is. And so the in short time and there were. Over 1.7 million Xan signatures and sold in this. So during that period I got to know a lot of health care workers and many of them and wrote to me and so it’s it’s kind of like a. Nice to see, like a little statistical analysis and could.
00:29:26.098 --> 00:29:27.460 help the community.
NOTE Confidence: 0.89303994
00:29:27.460 --> 00:29:30.428 And also in the spring on the so
NOTE Confidence: 0.89303994
00:29:30.428 --> 00:29:33.345 I did something that station are
NOTE Confidence: 0.89303994
00:29:33.345 --> 00:29:37.451 supposed to do and so that we spend
NOTE Confidence: 0.89303994
00:29:37.451 --> 00:29:40.594 quite a bit time how working with
NOTE Confidence: 0.89303994
00:29:40.594 --> 00:29:43.390 the state of Massachusetts and also
NOTE Confidence: 0.89303994
00:29:43.390 --> 00:29:46.234 with abroad and so helping shifting
NOTE Confidence: 0.89303994
00:29:46.234 --> 00:29:49.195 the PPE under swap on from China.
NOTE Confidence: 0.89303994
00:29:49.200 --> 00:29:52.854 And so I was on the state, Massachusetts.
NOTE Confidence: 0.89303994
00:29:52.854 --> 00:29:57.120 The task force in the spring and then like.
NOTE Confidence: 0.89303994
00:29:57.120 --> 00:29:59.226 One thing I was really touched
NOTE Confidence: 0.89303994
00:29:59.226 --> 00:30:00.279 last spring was.
NOTE Confidence: 0.89303994
00:30:00.280 --> 00:30:03.072 Many, many peoples and step in to help
NOTE Confidence: 0.89303994
00:30:03.072 --> 00:30:05.230 without asking expecting any credit.
NOTE Confidence: 0.89303994
00:30:05.230 --> 00:30:07.780 So they really a wonderful experience
NOTE Confidence: 0.89303994
00:30:07.780 --> 00:30:10.592 and by working with so many peoples
and who stepped in to help and so
like in the screen 'cause there
were not many flight from China to
US so was difficult to shift under
those medical supplies and two US
and then so was really wonderful.
Many people help out and so you can
see that there were four flight.
Shifting the usapyon swap watered
by the state of Massachusetts
from Shanghai to Boston under the
first was the flight of the first
flight leaving could own,
and because there were not many
commercial flight available
and travel flight available, so this flight was converted from the Air Canada Flight and the passenger flight to a charter flight and then the picture on the right is the first flight arriving Boston. And also I the innerspring or we launch how we feel up and so this app collects the information about the Covic 19 symptoms and behaviors and also testing a result. And so this was in collaboration with some junk. Many of you probably know fun by his work in CRISPR editing and gene editing and CRISPR and also
also banned Superman.

Who is the CEO of country? So this is really a great.

Collaboration between academia and industry 'cause we are not very good at developing up, but people in industry.

They're much better developing app.

So so many volunteer helping with this how we feel project and we build a nonprofit organization and with so many volunteers and then this app has over 750,000 users and also 50 million responses.
results and this is the first paper.

Out of this, how we feel project was published in Nature.

Human behavior last summer.

So here last spring,

who were more likely to be tested?

And it turns out that people

who had symptoms,

CDC symptoms or health care workers and

essential workers and people of color,

they were more likely to be tested.

So that makes sense,

because in the spring the testing

kids were not as widely available,

so the vulnerable group should

have priority to be tested.
And so this also present analysis challenge, because the people who were tested or likely to be sicker. And so therefore this is not a random sample, so when we studies Association between the factors associated with the infection, we have taken into account that people who were tested was not a random sample and so therefore in the analysis we use the inverse probability weighted procedures and to account for the selection bias. So we found that male with a higher risk of infection than females. And also we found that people of color.
were at higher risk of infection.

And also the essential workers and health care worker and these were at higher risk of infection.

Also, we found another household exposure and also community exposure are significant risk factor for infection and so you can see that for the household exposures after the show is almost 17 for Community exposures as we show almost three.

So what that mean is we need to break the within household and close place transmission and.

Cluding, the nursing home, homeless shelters and prisons,

homeless shelters and prisons,
and so and also we need to control the community transmission and so this finding was supported by the Massachusetts data that Massachusetts last year reported that almost 90% of covid cluster were household. So what that mean is household transmission is dominant is prevalent. Dominant lots of transmissions. And also we found the most important symptoms and was not the fever and cough was lots of peace and smell. So in particular we found out ratio is almost 33 associated with loss of taste and smell.
About 40% of those who were past positive had lost taste buds, and smell. Among those who are not testing about 6.6% among those who are test negative. That was about 5%. So this is an important symptom is also distinguished from the flu symptom. Then we also build a prediction model giving there were not enough tests available and then can we use the screening on two and two? Predict whether a person is likely to be infected or not. So by using the CDC symptom, you can see that the RC
NOTE Confidence: 0.8840181
00:35:16.450 --> 00:35:19.150 curve is AOC is about 70%.
NOTE Confidence: 0.8840181
00:35:19.150 --> 00:35:21.676 Using all the variables and it’s
NOTE Confidence: 0.8840181
00:35:21.676 --> 00:35:24.630 about 80% if we use a simpler
NOTE Confidence: 0.8840181
00:35:24.630 --> 00:35:26.740 model only used for variable,
NOTE Confidence: 0.8840181
00:35:26.740 --> 00:35:28.396 including the three exposure
NOTE Confidence: 0.8840181
00:35:28.396 --> 00:35:31.390 variable and also the loss of taste,
NOTE Confidence: 0.8840181
00:35:31.390 --> 00:35:31.845 smell,
NOTE Confidence: 0.8840181
00:35:31.845 --> 00:35:34.120 the symptom variables and then
NOTE Confidence: 0.8840181
00:35:34.120 --> 00:35:37.398 you can see the AOC is also 80%.
NOTE Confidence: 0.8840181
00:35:37.400 --> 00:35:39.776 And so this is very simple model but has
NOTE Confidence: 0.8840181
00:35:39.776 --> 00:35:42.099 very good predictability for infection.
NOTE Confidence: 0.8840181
00:35:42.100 --> 00:35:44.788 And when we build this model we use.
NOTE Confidence: 0.8840181
00:35:44.790 --> 00:35:48.342 This actually proves a boost that is a
NOTE Confidence: 0.8840181
00:35:48.342 --> 00:35:50.978 scalable gradient tree boosting method.
NOTE Confidence: 0.8840181
00:35:50.980 --> 00:35:54.193 So now let me talk about the
NOTE Confidence: 0.8840181
00:35:54.193 --> 00:35:56.740 defending challenge on the in 2021.
NOTE Confidence: 0.8840181
00:35:56.740 --> 00:36:00.276 So first is the vaccine rollout and optic,
NOTE Confidence: 0.8840181
00:36:00.280 --> 00:36:02.495 so the science was really
NOTE Confidence: 0.8840181
00:36:02.495 --> 00:36:03.824 wonderful last year,
NOTE Confidence: 0.8840181
00:36:03.830 --> 00:36:06.320 so developing the vaccine such as
NOTE Confidence: 0.8840181
00:36:06.320 --> 00:36:09.139 short time with such high efficacy,
NOTE Confidence: 0.8840181
00:36:09.140 --> 00:36:12.080 that’s really amazing and so.
NOTE Confidence: 0.8840181
00:36:12.080 --> 00:36:15.237 So the challenge is the vaccination program.
NOTE Confidence: 0.8840181
00:36:15.240 --> 00:36:16.148 So basically,
NOTE Confidence: 0.8840181
00:36:16.148 --> 00:36:20.200 how can we get the vaccine into people’s arm?
NOTE Confidence: 0.8840181
00:36:20.200 --> 00:36:23.350 And so that basically includes the
NOTE Confidence: 0.8840181
00:36:23.350 --> 00:36:26.230 distribution and also the administration.
NOTE Confidence: 0.8840181
00:36:26.230 --> 00:36:26.704 Also,
NOTE Confidence: 0.8840181
00:36:26.704 --> 00:36:29.074 it’s important to have equitable
NOTE Confidence: 0.8840181
00:36:29.074 --> 00:36:30.496 and scalable vaccination,
NOTE Confidence: 0.8840181
00:36:30.500 --> 00:36:32.865 and also is important to
NOTE Confidence: 0.8840181
00:36:32.865 --> 00:36:34.284 overcome vaccine hesitancy.
NOTE Confidence: 0.8840181
00:36:34.290 --> 00:36:38.686 I’m going to focus on this one.
NOTE Confidence: 0.77242935
00:36:38.690 --> 00:36:43.268 And the second defining challenge is
NOTE Confidence: 0.77242935
00:36:43.268 --> 00:36:47.260 the massive scalable testing and so.
NOTE Confidence: 0.77242935
00:36:47.260 --> 00:36:49.918 PCR test yes, a gold standard,
NOTE Confidence: 0.77242935
00:36:49.920 --> 00:36:52.811 but it is expensive and to do
NOTE Confidence: 0.77242935
00:36:52.810 --> 00:36:57.612 the massive regular testing.
NOTE Confidence: 0.77242935
00:36:57.612 --> 00:37:02.549 So I’m going to talk about
NOTE Confidence: 0.77242935
00:37:02.549 --> 00:37:04.541 efficient testing strategy using
NOTE Confidence: 0.77242935
00:37:04.550 --> 00:37:06.884 the pooled testing and also the
NOTE Confidence: 0.77242935
00:37:06.884 --> 00:37:10.417 other strategies rapid testing.
NOTE Confidence: 0.77242935
00:37:10.420 --> 00:37:13.268 And the third component is the
NOTE Confidence: 0.77242935
00:37:13.268 --> 00:37:17.174 implementation and compliance of
NOTE Confidence: 0.77242935
00:37:17.180 --> 00:37:19.612 public health control measures.
NOTE Confidence: 0.77242935
00:37:19.612 --> 00:37:20.187 So if you look at a quick job
NOTE Confidence: 0.77242935
00:37:13.268 --> 00:37:16.203 of the cases in January is not
NOTE Confidence: 0.77242935
00:37:16.203 --> 00:37:19.220 likely to do to the vaccine,
NOTE Confidence: 0.77242935
00:37:19.220 --> 00:37:22.146 because only less than 10% of the
NOTE Confidence: 0.77242935
00:37:22.146 --> 00:37:24.236 US population had been vaccinated.
NOTE Confidence: 0.77242935
00:37:24.240 --> 00:37:26.768 I seem like the last last month and
NOTE Confidence: 0.77242935
00:37:26.768 --> 00:37:29.063 the implementation and compliance
NOTE Confidence: 0.77242935
00:37:29.063 --> 00:37:31.708 and control measures and became
NOTE Confidence: 0.77242935
00:37:31.708 --> 00:37:34.549 better and people pay more attention
NOTE Confidence: 0.77242935
00:37:34.549 --> 00:37:37.105 to the behavior changes so that.
NOTE Confidence: 0.77242935
00:37:37.110 --> 00:37:40.550 Definitely is an important message.
NOTE Confidence: 0.77242935
00:37:40.550 --> 00:37:44.510 So let’s look at the vaccine rate and so.
NOTE Confidence: 0.77242935
00:37:44.510 --> 00:37:47.912 Overlap and this is from the one word data.
NOTE Confidence: 0.77242935
00:37:47.920 --> 00:37:49.111 You can see.
NOTE Confidence: 0.77242935
00:37:49.111 --> 00:37:51.493 Israel is definitely the role model
NOTE Confidence: 0.77242935
00:37:51.493 --> 00:37:54.430 and so the right now they have an
NOTE Confidence: 0.77242935
00:37:54.430 --> 00:37:57.019 average 70 doses and per 100 people.
00:37:57.020 --> 00:37:59.948 And so after we account that some people
have two doses on average about 40% people in Israel had been vaccinated
and with so that’s really amazing.
And you have this less than 10% and in particular,
you can see basically nobody in Africa has been vaccinated,
so that’s really not good.
00:38:02.700 --> 00:38:04.812 people in Israel had been vaccinated
and with so that’s really amazing.
And you have this less than 10%
00:38:02.700 --> 00:38:04.812 people in Israel had been vaccinated
and with so that’s really amazing.
And you have this less than 10%
00:38:04.812 --> 00:38:07.249 and with so that’s really amazing.
And you have this less than 10%
00:38:09.900 --> 00:38:12.436 if on the right you can see we
have a serious equity issue.
And in particular,
you can see basically nobody
00:38:14.960 --> 00:38:16.034 And in particular,
you can see basically nobody
00:38:16.034 --> 00:38:17.824 you can see basically nobody
00:38:17.824 --> 00:38:19.858 in Africa has been vaccinated,
00:38:19.860 --> 00:38:23.490 so that’s really not good.
00:38:19.860 --> 00:38:23.490 so that’s really not good.
00:38:23.490 --> 00:38:25.665 So the another defining challenges
Vaccine Induced Herd Immunity.
00:38:30.890 --> 00:38:33.500 we need to overcome vaccine hesitation.
NOTE Confidence: 0.77242935

00:38:33.500 --> 00:38:36.134 So I’m going to present the findings and from the how we feel
NOTE Confidence: 0.77242935

00:38:36.134 --> 00:38:39.246 data show McCabe is my Postal.
NOTE Confidence: 0.77242935

00:38:39.246 --> 00:38:41.760 He take a lead in this work in collaboration with many colleagues.
NOTE Confidence: 0.77242935

00:38:41.760 --> 00:38:45.424 So here is a way, a lunch, the Maxim question in how we fill up in early December.
NOTE Confidence: 0.77242935

00:38:45.424 --> 00:38:47.930 So with here the result of analyzing the first month data about 30,000 people.
NOTE Confidence: 0.77242935

00:38:50.728 --> 00:38:52.972 fill up in early December.
NOTE Confidence: 0.77242935

00:38:55.310 --> 00:38:58.366 So with here the result of analyzing the first month data about 30,000 people.
NOTE Confidence: 0.77242935

00:39:01.050 --> 00:39:03.684 So we develop a partnership with Kinetica and last spring and so Kinetica and last spring and so
NOTE Confidence: 0.77242935

00:39:06.166 --> 00:39:08.750 that’s why you can see we have more respondents and in the kinitica,
NOTE Confidence: 0.77242935

00:39:11.300 --> 00:39:13.350 and also because the countries
00:39:13.350 --> 00:39:14.990 is located in California.

00:39:14.990 --> 00:39:18.206 So we had more respondent respondent.

00:39:18.210 --> 00:39:19.132 In California, so if you look at overall vaccine hesitancy,

00:39:19.132 --> 00:39:22.820 hesitancy read,

00:39:22.820 --> 00:39:23.742 you can see like thoughts are more hesitant,

00:39:23.742 --> 00:39:27.430 and so overall the vaccine hasn’t

00:39:27.430 --> 00:39:30.178 hesitancy rate is about 1818% from the

00:39:30.178 --> 00:39:33.422 hallway field data and 82% on the.

00:39:33.422 --> 00:39:36.632 What said they are likely were

00:39:36.632 --> 00:39:39.338 more likely to take the vaccine.

00:39:39.338 --> 00:39:42.138 So if you look hard, um,

00:39:42.140 --> 00:39:44.490 vaccine hesitancy rate by race and ethnicity,

00:39:44.490 --> 00:39:47.220 then you can see that people of color

00:39:47.220 --> 00:39:50.028 are much more likely to be vaccine hesitant.
So in particular, if you look at a black for example, the vaccine hesitancy is all. 46%, almost 50% so so compared to white is about 15%, but compared to Hispanic, about 30% you can see a large fraction of them are undecided group. So what that mean is that a community engagement through the education of outreach is important. Who are more likely to be vaccine hesitant and so we found the younger people are more likely to
be a vaccine hesitant and females, and also health care worker essential workers and also the people of color. In particular, black people are 3.5 times more likely to be vaccine hesitant than white, and people with pre existing conditions and low income. And also rural areas and also the thoughts and also places with high kufic burden and also the people who they are. So those are more likely to be vaccine hesitant people who wear masks and also use the protective measures. They are less likely to be vaccine hesitant.
Talk to us in summary.

So the vulnerable group are more likely to be vaccine hesitant, and they include people of color, healthcare workers, essential workers, young people, female, parents, and low-income people. Also people not using protective measures.

An Irish last year, the lunch, a community engagement alliance. So an Irish last late last year, the lunch, a community engagement alliance.
involved multiple centers and the one of the goal is to do the Community engagement to help with participation. Overcome the vaccine hesitancy. So what this tells us is community engagement for vaccination, of which an education is important, so that Pic home number 5 is important, remained bigil and to scale up scale control measure and vaccination up the control measure and vaccination by protecting the vulnerable group, including the health care workers and essential workers and elderly. And also it’s important.
00:42:39.228 --> 00:42:41.538 to reach the zero kovik.
NOTE Confidence: 0.8238236
00:42:41.540 --> 00:42:43.940 So what that mean is.
NOTE Confidence: 0.8238236
00:42:43.940 --> 00:42:46.404 We need to be careful and reopen
NOTE Confidence: 0.8238236
00:42:46.404 --> 00:42:49.175 slowly when the number of cases are
NOTE Confidence: 0.8238236
00:42:49.175 --> 00:42:51.225 sufficiently small and also with
NOTE Confidence: 0.8238236
00:42:51.225 --> 00:42:53.962 the control measures are so if when
NOTE Confidence: 0.8238236
00:42:53.962 --> 00:42:56.657 we opened too early and we slipped
NOTE Confidence: 0.8238236
00:42:56.657 --> 00:42:59.051 in the control measure like what
NOTE Confidence: 0.8238236
00:42:59.051 --> 00:43:01.600 happened last summer and in the South,
NOTE Confidence: 0.8238236
00:43:01.600 --> 00:43:04.071 and is likely to see the researchers
NOTE Confidence: 0.8238236
00:43:04.071 --> 00:43:06.449 and also is important to pay
NOTE Confidence: 0.8238236
00:43:06.449 --> 00:43:08.519 attention to the long color,
NOTE Confidence: 0.8238236
00:43:08.520 --> 00:43:11.656 the long term effect especially among the
NOTE Confidence: 0.8238236
00:43:11.656 --> 00:43:15.017 young people and then also the to a build.
NOTE Confidence: 0.8238236
00:43:15.020 --> 00:43:17.948 I’ve seen uptick and it’s important
NOTE Confidence: 0.8238236
00:43:17.948 --> 00:43:20.440 to have community engagement and
outreach and build public trust.

So basically, how can we implement the?

Control measures and also implement vaccination and ensure high compliance is the defining challenge this year.

And the truth, the other component is for this year is how can we boost the testing capacity and buy a cover by doing more test.

And so because it’s uh if one needs to do the test frequently and to do the PCR test is difficult to do for many institution because it’s costly and so they put the testing provide an alternative.
So I'm going to talk about this efficient put testing. A design using the hyper graph factorization first. What is the protesting? The goal is that would put testing is to screen a large population with a few tests and giving the limited resources. So this will help reopen the school safely and the simple idea is used. This uh document design sofa. Suppose we have 100 people and we do 20 tests and then so we pulled the people sample into different. Pools and suppose there's only one case, and then we test each pool support.
Each pool has a 10 people. And then we tested each put do 10 pull test and how we found this cool is costing. Then we test every individual in this pool so in. Therefore instead of doing 100 test you only do 20 tests and so this is the basic idea. Put testing. So what is the limitation of this simple of protesting design? And so this document design allow one individual go to one pool that is Q equal to 1, then cycle through the pool until
all individuals are assigned. So if you look at this example with eight subjects and six pool, then you can see that we assign the face first six subject to the segments and each subject. And do the pull A&B so only one person per pool, so this is not optimal, only one pool, only one pool per person and this could lead to a non redundancy could lead to a non reduc
cancy that also reduce the sensitivity. So the question is can we do better? Can we assign each individual to more than one pool?
That basically makes a Q equal to two.

So let’s start from something like if there’s a safe assign one person to two pools.

So for example I assigned the first person to. the first person to.

2A B second person to put a C and third person to pull busy and so on and then cycle through the order.

So that basically this idea assign each person to two pools.

What is the problem?

The problem is by doing this simple way the design is not balanced.

You can see that pull it has
five subjects and puppy has four
and pull up as only one subject.
Because when does them the
pulling and by assigning one
person to more than one pool
while need to dilute the sample.
So if one has a different solution for
different pools that will affect the.
Accuracy under then the sensitivity.
So can we do better so that is
a basic idea of a more balanced
design we call the hyper design.
So this using the hyper graph factorization.
So the basic idea is we want to
make the spell is that possible?
So for example like here you can see
00:47:06.772 --> 00:47:09.774 that assigned person A to pull a BE

00:47:09.774 --> 00:47:12.365 person B person to pull CD person

00:47:12.365 --> 00:47:15.402 3 two pull ENF person four to pull.

00:47:15.402 --> 00:47:18.930 PNC Person 5 to pull the D&F and so

00:47:19.026 --> 00:47:22.370 on and so this. This idea is after you.

00:47:22.370 --> 00:47:24.590 Each pool has four samples and

00:47:24.590 --> 00:47:27.696 so you can see for the 1st pool

00:47:27.696 --> 00:47:29.616 and the test is negative.

00:47:29.620 --> 00:47:32.294 The second pull the test is positive.

00:47:32.300 --> 00:47:34.967 3rd pool passes positive and so on.

00:47:34.970 --> 00:47:37.763 Then afterwards we do the pool testing

00:47:37.763 --> 00:47:40.960 and then we can decode to see that

00:47:40.960 --> 00:47:43.759 which person is likely to be a case.

00:47:43.760 --> 00:47:46.388 And here you can see that.

00:47:46.390 --> 00:47:49.330 After do the decoding person 3, four,
and seven are likely to be a positive, and then we test each of them individually and find out persons. Seven are the cases and so why it is called hyper graph design. And that because this is related to the hypergraph, and in complete awe metrics one will assign this. This is edge person one and so that's assigned the pool at A&B and then person two assigned to
C&D and so so this is so that’s why it’s called a hypergraph. So basically what we do is we need to assign the individuals and in the right sequence to make them as balanced as possible and not overlap as much as possible. And so by doing this design, when we kill equal to 216 pool, we have 5 factorizations and we have 5 factorizations and so you can see for each factorization there’s no overlap and also between every two consecutive. Assignment under then there is no overlap as well.
And by doing this hypergraph designs and so you can see that we can have a balanced pool and also is very easy to implement and so this Calculator and so and also very easy to decode and so this is based on the company Atomic Comics. Population so we can do this calculations and for Q equal two and three, but for Q equals greater than three, the calculations much more challenging. And so by doing that then you can see that. And here we plot out the. Efficiency against the prevalence, so only if the prevalence is low is worthwhile to do.
NOTE Confidence: 0.8389234
00:49:50.970 --> 00:49:53.476 Put testing if the prevalence is high, there's no need to do put testing so you can see that doing the hyper design and it is. Efficient and then the efficiency is almost 6 compared to individual design and also expect her than a redesign. That efficiency is 4.6 and when the preference become higher and then you can see that the efficiency goes down and comparing the hyper design with a radius and efficient, the sensitivity is pretty similar and also when we have 384 subject per batch also when you can see that the hyper design still outperformed the other design.
And for the Pytest peoples design.

And so it.

Especially when the prevalence become higher and then you can see that uh, sensitivity almost reach to 0.

And so this also thought we look at a different design in different scenarios, and we showed that is hyper design.

Suppose each day we collect 3000 samples.

Then you can see that efficiency screening
capacity using this Q equal to two is 122.

That is much.

A better and so so then also

if one has a Q equal to three, that means a law allowing assigning one person to three pools.

Then in that situation we need to use the hypergraph and with the.

Those kind of 20 different hyperedges.

So in summary, to scale up widespread testing, hyper this is based on hypergraph factor factorization design provide efficient pool design to maximize the balance and efficiency,
and the protesting is useful when the prevalence is low when the preferences highs,
there’s no need for protesting,
just do the individual testing and we build a website that allows the investigator and the two design their own study.
And so to combat kovik and so we are really in this together.
And so we have to be together and be stronger.
And so it’s important to let the data speak and also develop evidence based strategy.
And we show that there are two feature of the Covic. One is is highly transmissible, second is highly convert.
And also it’s important to remain vigilant and to use the multifaceted interventions. And to combat Covid, there are multiple defining challenges this year. One is a Black Max magazine, distribution, uptake and education. The other is a scalable testing, so we talk about the put testing and so I want to thank the many of the collaborators and so who made many contributions to help with the project. And also there’s a quick announcement and the cops and less at lunch. This Covid 19 data.
Lisa W eaponer last December and so this is every two weeks and on Thursday from 12:00 to 1:00.

And so these are we have the last few two month. Last month. We have a wonderful speaker. We have a wonderful speaker. Great turn out.

So those are the speaker in the coming weeks and from Denmark Mukherjee who will talk about Covic. Many of you know and he’s a former president of National Academy of Medicine and also.

Jim Young Kim is a former president of World Bank and so they’re going to
give the next week talks and thank you.

Thanks young for this wonderful talk is very useful.

I want to weather the audience have any questions for she home.

Yeah I have a question.

So I’m wondering,

will people who are willing to respond to the how we feel study be more likely to have lower hesitancy?

I would think the how we feel.

We study people probably.

I would think that

probably likely to be true,
and so the how we feel samples because of people use the app, so at least that they are coping aware and they think a quickly is problem and so it’s possible that in the national samples when we have a more representative samples and the hesitancy rate may be higher.

Donna has a question. Yeah, I see her name was incredible work you’ve done. It’s just absolutely phenomenal and breathtaking. How you’ve addressed each issue arising in the kobid epidemic, one by one and come up with such clarity.
To guide us. So my question is about the hyper designs. I’ve been aware of pool testing, which I know. We all know it’s been around for awhile, but I’m just wondering, you know, is there like a rule of thumb like safe the prevalence rate is like 5%? How many digit in your graph like how many fewer tests would you have to use using a hyper design versus like the standard approach that you know people would tend to use which is to just test everybody. Yeah, so that is if you can see that from here.
Yeah, it’s a little hard to see it’s a little small. Oh this hyper yeah I can the the efficiency so you can see the efficiency that is about hyper design. Yes almost six so that means that we can each task and have 6 people and by individual design so you can see that suppose you have 100 people. This is 96 so 96 / 6 and then then you can see that that is. How many? Yes, I think about it. If you do individual test that is 100. What is the?
I didn’t understand really what the 96 and the 3:50.

Yeah batch, so there are 96 so in the so if you think about when you win.

If you think about it already.

You have to think about.

The Matrix is 8 by 12 matrix.

You put all the samples and in this 8 by 12 array.

OK, thank you yeah.

And also if you look at the capacity.
here you can see the capacity is much better so you can see that.

Suppose I need to test 3000 people a day. Execution can only afford half 12 tests and then you can see the efficiency.

Screening capacity is almost 120. So that’s really good. That’s very, very good, yeah?

You still have a question or say thank you very much. OK, so I also have a related question. So she how you mentioned that before. A future work you want to perform a regarding the reproduction number estimation and this intervention
work is to consider different other covariates when you are modeling the reproduction rate. So I wonder, have you also considered like trying to take into consideration different type of various the mutation of different various and then maybe certain high reproduction rate? And perhaps others. This process? Yeah, that is excellent suggestions on the so yeah, if we could have those data will be great that we could include those in the model.
00:58:35.784 --> 00:58:38.488 besides the different type of variance.
NOTE Confidence: 0.82796884
00:58:38.490 --> 00:58:40.950 And also like the vaccination rate.
NOTE Confidence: 0.82796884
00:58:40.950 --> 00:58:44.214 That would be a very good variable included.
NOTE Confidence: 0.82796884
00:58:44.220 --> 00:58:47.892 And so the challenge for us right now is,
NOTE Confidence: 0.82796884
00:58:47.900 --> 00:58:50.441 as you know UK has been doing
NOTE Confidence: 0.82796884
00:58:50.441 --> 00:58:53.220 a great job in the sequencing,
NOTE Confidence: 0.82796884
00:58:53.220 --> 00:58:56.076 viral sequencing and so in other words the
NOTE Confidence: 0.82796884
00:58:56.076 --> 00:58:58.310 surveillance and sequencing surveillance.
NOTE Confidence: 0.82796884
00:58:58.310 --> 00:59:02.107 But not US, and so we have not doing a great
NOTE Confidence: 0.82796884
00:59:02.107 --> 00:59:05.555 job in sequencing and so so therefore the
NOTE Confidence: 0.82796884
00:59:05.555 --> 00:59:09.269 UK could monitor the new virus and well,
NOTE Confidence: 0.82796884
00:59:09.270 --> 00:59:12.204 but I think with the one of the things
NOTE Confidence: 0.82796884
00:59:12.204 --> 00:59:15.606 we need to do this year is to increase
NOTE Confidence: 0.82796884
00:59:15.606 --> 00:59:17.938 the various viral sequencing capacity
NOTE Confidence: 0.82796884
00:59:17.938 --> 00:59:21.445 so we could monitor the new variants.
NOTE Confidence: 0.82796884
00:59:21.450 --> 00:59:24.915 So then also make the data available and to
the public and then that can be included. In the analysis. So what I've found that last year and during the Covid people were much more willing to share the data computer for an, though this is really, really wonderful and also the much more preprint and compared to before, more preprint and compared to before, and that were posted in about archive and made archive, and people were really willing and to share their findings to the Community as soon as possible. So these are really wonderful spirit about open science and.
And is.

Fired on the by many researchers last year.

Thanks, that’s really informed him.

Sorry, go ahead who is trying to ask a question.

Me, I mean yeah, I see who I have a question regarding to this pulling test. You said the pulling test and individual mix with five other. Cure lung disease samples.

Basically, the concentration is diluted,
01:00:49.938 --> 01:00:53.990 so how does sensitivity will be keep
NOTE Confidence: 0.7797129
01:00:53.990 --> 01:00:58.610 the same and how the next ways how
NOTE Confidence: 0.7797129
01:00:58.610 --> 01:01:02.080 to compare if there’s one positive
NOTE Confidence: 0.7797129
01:01:02.080 --> 01:01:04.968 case with five individual their
NOTE Confidence: 0.7797129
01:01:04.968 --> 01:01:08.438 normal cases an A normal situation
NOTE Confidence: 0.7797129
01:01:08.438 --> 01:01:11.906 controls and compare with all six.
NOTE Confidence: 0.7797129
01:01:11.910 --> 01:01:14.222 There just get exposure
NOTE Confidence: 0.7797129
01:01:14.222 --> 01:01:16.250 with low concentration.
NOTE Confidence: 0.7797129
01:01:16.250 --> 01:01:19.410 So there will be probably have
NOTE Confidence: 0.8333051
01:01:19.410 --> 01:01:21.522 some sensitivity issue if
NOTE Confidence: 0.8333051
01:01:21.522 --> 01:01:23.630 pulling together compared to
NOTE Confidence: 0.8333051
01:01:23.630 --> 01:01:26.792 individual tests, then the error for
NOTE Confidence: 0.8333051
01:01:26.792 --> 01:01:29.424 measurement error testing error for
NOTE Confidence: 0.8333051
01:01:29.424 --> 01:01:32.586 the two different types of tests.
NOTE Confidence: 0.8333051
01:01:32.586 --> 01:01:35.748 How did you consider them additional
NOTE Confidence: 0.8333051
01:01:35.750 --> 01:01:38.378 to the hyper structured testing?
NOTE Confidence: 0.8514964
01:01:39.020 --> 01:01:41.620 Yes, I think this is a great question.
NOTE Confidence: 0.8514964
01:01:41.620 --> 01:01:43.895 Sorry I did not make that clear.
NOTE Confidence: 0.8514964
01:01:43.900 --> 01:01:46.537 What I meant was that hyper design and are
NOTE Confidence: 0.8514964
01:01:46.537 --> 01:01:49.100 ready that they had a similar sensitivity,
NOTE Confidence: 0.8514964
01:01:49.100 --> 01:01:50.720 but the sensitivity is lower
NOTE Confidence: 0.8514964
01:01:50.720 --> 01:01:52.016 than the individual tests.
NOTE Confidence: 0.8514964
01:01:52.020 --> 01:01:56.820 If you look at the curve in about I see.
NOTE Confidence: 0.8514964
01:01:56.820 --> 01:01:59.116 At the green and red they have
NOTE Confidence: 0.8514964
01:01:59.116 --> 01:02:00.869 a similar sensitivity by the
NOTE Confidence: 0.8514964
01:02:00.869 --> 01:02:02.624 compared to the individual test.
NOTE Confidence: 0.8514964
01:02:02.630 --> 01:02:05.446 That is, this black line and it has
NOTE Confidence: 0.8514964
01:02:05.446 --> 01:02:07.504 higher sensitivity and so then they
NOTE Confidence: 0.8514964
01:02:07.504 --> 01:02:09.831 as you are definitely right when one
NOTE Confidence: 0.8514964
01:02:09.831 --> 01:02:12.204 do one month does the pooled testing
NOTE Confidence: 0.8514964
01:02:12.204 --> 01:02:14.600 because the sample needs to be diluted,
so therefore we need to pay a price
and then sensitivity will be lower.
Yeah, so the overall one the population.

Pilots large scale testing.
We may have more undetectable test.
Think about if six samples always happen
is 1 sample has a positive case,
so we may have some testing error
here. Yeah yeah, so yes.

Can I just jump in for a second?
So 'cause I think the comparison as
you could compare this hyper design
you could compare this hyper design to just testing everybody or the
hyper design to the traditional
01:02:56.055 --> 01:02:58.130 pool testing approach where you
NOTE Confidence: 0.852674794
01:02:58.205 --> 01:03:00.144 just divide 100 people in each one
NOTE Confidence: 0.852674794
01:03:00.144 --> 01:03:02.860 is in a single batch and I think
NOTE Confidence: 0.852674794
01:03:02.860 --> 01:03:05.343 she hung what you’re saying and it
NOTE Confidence: 0.852674794
01:03:05.343 --> 01:03:07.825 makes sense to me intuitively is by
NOTE Confidence: 0.852674794
01:03:07.825 --> 01:03:09.600 repeating people in multiple batches
NOTE Confidence: 0.852674794
01:03:09.600 --> 01:03:11.730 were increasing the chances of having
NOTE Confidence: 0.852674794
01:03:11.730 --> 01:03:14.362 doubles and triples in the same batch.
NOTE Confidence: 0.852674794
01:03:14.362 --> 01:03:17.344 And then lowering the chance of having
NOTE Confidence: 0.852674794
01:03:17.344 --> 01:03:20.035 false negatives as opposed to the
NOTE Confidence: 0.852674794
01:03:20.035 --> 01:03:22.925 traditional design where you take the
NOTE Confidence: 0.852674794
01:03:22.925 --> 01:03:26.698 100 people and they’re only in one match.
NOTE Confidence: 0.852674794
01:03:26.700 --> 01:03:27.804 Is that right?
NOTE Confidence: 0.852674794
01:03:27.804 --> 01:03:29.644 Yeah, that’s so that’s right.
NOTE Confidence: 0.852674794
01:03:29.650 --> 01:03:32.212 So you don’t want to put a
NOTE Confidence: 0.852674794
01:03:32.212 --> 01:03:34.448 hundred 100 people in one batch,
because if you do that, then the sample need to diluted a lot and then you will sacrifice the sensitivity. So that’s why when want to do the optimal design, want to account for both the balance and also sensitivity and including both of them and so then. So that’s why when we build this when we define this efficient screening capacity, this calculation. That incorporated sensitivity in the calculation as well.
01:04:06.640 --> 01:04:09.736 So if you look at the traditional design,
NOTE Confidence: 0.8346741
01:04:09.740 --> 01:04:12.844 so here you can see that each person,
NOTE Confidence: 0.8346741
01:04:12.850 --> 01:04:14.071 the traditional design.
NOTE Confidence: 0.8346741
01:04:14.071 --> 01:04:16.106 Basically each person is assigned
NOTE Confidence: 0.8346741
01:04:16.106 --> 01:04:18.798 to a single pool and so this
NOTE Confidence: 0.8346741
01:04:18.798 --> 01:04:20.603 is this called document design,
NOTE Confidence: 0.8346741
01:04:20.610 --> 01:04:23.886 and so this design so you can
NOTE Confidence: 0.8346741
01:04:23.886 --> 01:04:27.529 see that the six for six people.
NOTE Confidence: 0.8346741
01:04:27.530 --> 01:04:32.100 In this example, like the person, one and.
NOTE Confidence: 0.8346741
01:04:32.100 --> 01:04:35.076 Assigned to a person to assign to P,
NOTE Confidence: 0.8346741
01:04:35.080 --> 01:04:37.698 and so this is not efficient design.
NOTE Confidence: 0.8346741
01:04:37.700 --> 01:04:40.276 And so if we assign each person to
NOTE Confidence: 0.8346741
01:04:40.276 --> 01:04:42.270 multiple pools and after decoding
NOTE Confidence: 0.8346741
01:04:42.270 --> 01:04:44.405 that will improve the efficiency.
NOTE Confidence: 0.80364835
01:04:46.760 --> 01:04:48.375 So generally the cute does
NOTE Confidence: 0.80364835
01:04:48.375 --> 01:04:49.990 should not be too big.
So here you can secure equal to 1.
That means one person assigned to one pool equals 2 means a person sent to two pools and just think about this is very interesting and so you can see the basically using the graph and the ABCD basically means the pool and each edge indicated person. So you can see this person one and then this person too. So this person too. And then so on.
01:05:23.850 --> 01:05:25.740 Thanks young, I have one.
NOTE Confidence: 0.78496486
01:05:25.740 --> 01:05:27.620 I have one last question.
NOTE Confidence: 0.78496486
01:05:27.620 --> 01:05:29.606 If other people do not have
NOTE Confidence: 0.78496486
01:05:29.606 --> 01:05:31.869 more question so I wonder is
NOTE Confidence: 0.78496486
01:05:31.869 --> 01:05:34.029 also related to the sensitivity.
NOTE Confidence: 0.78496486
01:05:34.030 --> 01:05:36.284 I wonder how we considered too instead
NOTE Confidence: 0.78496486
01:05:36.284 --> 01:05:39.017 of using on the testing directly but
NOTE Confidence: 0.78496486
01:05:39.017 --> 01:05:41.567 construct some posterior for each person.
NOTE Confidence: 0.78496486
01:05:41.570 --> 01:05:43.078 Use other covariates including
NOTE Confidence: 0.78496486
01:05:43.078 --> 01:05:44.209 your past history.
NOTE Confidence: 0.78496486
01:05:44.210 --> 01:05:48.056 Whether you have higher risk an.
NOTE Confidence: 0.78496486
01:05:48.060 --> 01:05:50.424 I wonder if we use such
NOTE Confidence: 0.78496486
01:05:50.424 --> 01:05:51.606 personalized information combined
NOTE Confidence: 0.78496486
01:05:51.606 --> 01:05:53.400 with this testing results,
NOTE Confidence: 0.78496486
01:05:53.400 --> 01:05:55.460 can we have better sensitivity?
NOTE Confidence: 0.8241894
01:05:56.990 --> 01:05:58.592 Very good question.
Yeah, I can see the potential.

I can see the potential for doing that.

Yeah, I think they’re right now in the screening program and the no.

Demographic information is collected, and so only the sample collected.

So for example, like abroad they.

But bro, the dead on.

The spring when we first started it was about maybe a 3000 of sample per day and so right now.

As you know,

broad cover almost 9025% of the testing,

and in the New England areas almost cover like 3,000,000 and test and so
the event of data were lots of data.

Sand were collected and.

In the testing,

but those data cannot be used for research.

Thanks, I don’t know if the audience have further question.

Maybe you can also email see how afterwards.

So we’re running a little bit overtime,

but it’s very. This wonderful talk.

Can we have learned so much

from Seahawks services talk?

Thank you again. Thank you very much.