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 everyone's 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 Covic 19 so this is the data just downloaded earlier this week. 
So as you can see that the right now
00:01:56.312 --> 00:01:58.662 they’re over 110,000,000 cases in

00:01:58.662 --> 00:02:02.386 US and also 2.4 two point 4,000,000

00:02:02.386 --> 00:02:05.081 deaths were 110,000,000 cases and

00:02:05.090 --> 00:02:06.582 2.4 million deaths worldwide.

00:02:06.582 --> 00:02:10.217 So if you look at the curve on the left,

00:02:10.220 --> 00:02:12.103 that is the case curve and for

00:02:12.103 --> 00:02:13.560 a few selective countries.

00:02:13.560 --> 00:02:15.970 So as you can see for both the UK and

00:02:16.037 --> 00:02:18.361 also United States and the number of

00:02:18.361 --> 00:02:20.858 cases had been going down in January,

00:02:20.860 --> 00:02:22.678 so that is a good sign.

00:02:22.680 --> 00:02:24.724 And also Israel as you know that

00:02:24.724 --> 00:02:26.901 Israel has been a really leader in

00:02:26.901 --> 00:02:28.761 a vaccination and so their cases

00:02:28.825 --> 00:02:30.589 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 columnist, he is currently a 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 countries 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

deided 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 and technology 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 so you can see there’s a quite few coverage of the findings and 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

NOTE Confidence: 0.83434194
00:07:35.411 --> 00:07:38.135 members like a senators on it.
NOTE Confidence: 0.83434194
00:07:38.140 --> 00:07:40.975 And then they later on they also
NOTE Confidence: 0.83434194
00:07:40.975 --> 00:07:43.090 invited few other witnesses.
NOTE Confidence: 0.83434194
00:07:43.090 --> 00:07:46.434 And then they wrote a letter to the
NOTE Confidence: 0.83434194
00:07:46.434 --> 00:07:49.327 Prime Minister Johnson and they make
NOTE Confidence: 0.83434194
00:07:49.327 --> 00:07:52.285 a 10 recommendations in their letter.
NOTE Confidence: 0.83434194
00:07:52.290 --> 00:07:55.394 Under so so then I I was honored
NOTE Confidence: 0.83434194
00:07:55.394 --> 00:08:00.602 that several of my recommendation
NOTE Confidence: 0.83434194
00:08:00.602 --> 00:08:03.987 we included in their recommendation
NOTE Confidence: 0.83434194
00:08:03.987 --> 00:08:07.299 to the Prime Prime Minister Johnson.
NOTE Confidence: 0.83434194
00:08:07.300 --> 00:08:08.392 And then, uh,
NOTE Confidence: 0.83434194
00:08:08.392 --> 00:08:11.597 this last month and as a one year
NOTE Confidence: 0.83434194
00:08:11.597 --> 00:08:14.981 anniversary of the colic outbreak and
NOTE Confidence: 0.83434194
00:08:14.981 --> 00:08:17.869 nature published their third editorial
NOTE Confidence: 0.83434194
00:08:17.869 --> 00:08:21.187 and 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 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.
NOTE Confidence: 0.845761500000001
00:09:34.012 --> 00:09:37.244 on the in this orphan and seafood
NOTE Confidence: 0.845761500000001
00:09:37.244 --> 00:09:39.504 market and this seafood market
NOTE Confidence: 0.845761500000001
00:09:39.587 --> 00:09:41.647 was closed on January 1st.
NOTE Confidence: 0.845761500000001
00:09:41.650 --> 00:09:44.786 Under then on March on January 23rd,
NOTE Confidence: 0.845761500000001
00:09:44.790 --> 00:09:50.370 and that was a lunch of the lock down and so.
NOTE Confidence: 0.845761500000001
00:09:50.370 --> 00:09:53.592 As you can see that before the lock down
NOTE Confidence: 0.845761500000001
00:09:53.592 --> 00:09:56.799 the estimated RTT values is about 3.5,
NOTE Confidence: 0.845761500000001
00:09:56.800 --> 00:10:00.104 so that means each purse case could be.
NOTE Confidence: 0.845761500000001
00:10:00.110 --> 00:10:03.809 Like 3.5 people and so that is not good.
NOTE Confidence: 0.845761500000001
00:10:03.810 --> 00:10:06.393 So that means this is really the
NOTE Confidence: 0.845761500000001
00:10:06.393 --> 00:10:08.009 disease is very transmissible
NOTE Confidence: 0.845761500000001
00:10:08.009 --> 00:10:10.970 and then after the lock down and
NOTE Confidence: 0.845761500000001
00:10:10.970 --> 00:10:13.171 with the social distancing is
NOTE Confidence: 0.845761500000001
00:10:13.171 --> 00:10:16.132 really how the RT value dropped to
NOTE Confidence: 0.845761500000001
00:10:16.140 --> 00:10:18.606 a little over 1.2 and then.
NOTE Confidence: 0.845761500000001
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 onsite. 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 symptomatic period so this period, the person, even though a person doesn’t have
a symptom and but very load,  
NOTE Confidence: 0.8317293  
00:11:44.440 --> 00:11:48.066 is high enough and then the person  
NOTE Confidence: 0.8317293  
00:11:48.066 --> 00:11:49.102 could become.  
NOTE Confidence: 0.8317293  
00:11:49.110 --> 00:11:51.042 Infectious and so this period is  
NOTE Confidence: 0.8317293  
00:11:51.042 --> 00:11:52.852 about two days between presymptomatic  
NOTE Confidence: 0.8317293  
00:11:52.852 --> 00:11:54.680 two symptomatic and between  
NOTE Confidence: 0.8317293  
00:11:54.680 --> 00:11:56.965 exposure to pre symptomatic period.  
NOTE Confidence: 0.8317293  
00:11:56.970 --> 00:12:01.558 This is about 3 days,  
NOTE Confidence: 0.8317293  
00:12:01.558 --> 00:12:03.260 and so we built in those components  
NOTE Confidence: 0.8317293  
00:12:03.260 --> 00:12:05.220 and in the model,  
NOTE Confidence: 0.8317293  
00:12:05.220 --> 00:12:06.004 presymptomatic compartment,  
NOTE Confidence: 0.8317293  
00:12:06.010 --> 00:12:08.882 and also because at that time the testing  
NOTE Confidence: 0.8317293  
00:12:08.882 --> 00:12:11.509 case will not that widely available,  
NOTE Confidence: 0.8317293  
00:12:11.510 --> 00:12:13.988 and so therefore there were a lot  
NOTE Confidence: 0.8317293  
00:12:13.988 --> 00:12:16.230 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 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. 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 sudden after supposed when user, social distancing and lock down social distancing and lock down number of cases still going up number of cases still going up 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 two 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 of the 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
00:16:33.919 --> 00:16:35.999 take home take home messages,
NOTE Confidence: 0.8223038
00:16:36.000 --> 00:16:37.905 the social distancing and centralized
NOTE Confidence: 0.8223038
00:16:37.905 --> 00:16:39.429 isolation quarantine were critical
NOTE Confidence: 0.8223038
00:16:39.429 --> 00:16:41.119 for controlling the outbreak,
NOTE Confidence: 0.8223038
00:16:41.120 --> 00:16:43.215 so using the social distancing
NOTE Confidence: 0.8223038
00:16:43.215 --> 00:16:46.221 alone that help but was not good
NOTE Confidence: 0.8223038
00:16:46.221 --> 00:16:48.693 enough so that helped make our
NOTE Confidence: 0.8223038
00:16:48.693 --> 00:16:50.932 reduce around 1:00 and but did
NOTE Confidence: 0.8223038
00:16:50.932 --> 00:16:53.457 not bend the curve and the reason
NOTE Confidence: 0.8223038
00:16:53.457 --> 00:16:57.076 is the there were lots of the.
NOTE Confidence: 0.8223038
00:16:57.080 --> 00:16:58.892 Community transmissions on the
NOTE Confidence: 0.8223038
00:16:58.892 --> 00:17:01.157 social distancing help block the
NOTE Confidence: 0.8223038
00:17:01.157 --> 00:17:03.207 community transmission that is
NOTE Confidence: 0.8223038
00:17:03.207 --> 00:17:05.243 between household transmission but
NOTE Confidence: 0.8223038
00:17:05.243 --> 00:17:07.400 within family transmission and cost.
NOTE Confidence: 0.8223038
00:17:07.400 --> 00:17:09.276 Place transmission with common
and so 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.

So you can see that Italy also did
the social distancing that help reduce

the R and the R curve lingered around one for over a month and did not bend

that Germany data in the spring.

The same thing under the curve did not band.

And the second feature from analyzing the Wuhan data is the

Covic is highly converged on the, so we estimated about 87% of the cases were undetected.

So in other words, the detected cases was only the tip of the iceberg,

and so you can see that on the on

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

entertainment rate and so we estimated

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,

that was about 3%.

And then US result very similar.

the CDC did theological study last
year 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. When 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
NOTE Confidence: 0.84831905
00:19:46.636 --> 00:19:48.687 days by lifting all the control
NOTE Confidence: 0.84831905
00:19:48.687 --> 00:19:50.865 measures and the first strategy is
NOTE Confidence: 0.84831905
00:19:50.865 --> 00:19:53.109 after the first day observing the.
NOTE Confidence: 0.84831905
00:19:53.110 --> 00:19:55.200 Zero confirmed Case No matter
NOTE Confidence: 0.84831905
00:19:55.200 --> 00:19:58.034 whether the second day has the case
NOTE Confidence: 0.84831905
00:19:58.034 --> 00:20:00.358 or not and when to reopen info.
NOTE Confidence: 0.84831905
00:20:00.360 --> 00:20:03.896 This and the second strategy is one has
NOTE Confidence: 0.84831905
00:20:03.896 --> 00:20:07.559 a confirmed 0 case for 14 consecutive days.
NOTE Confidence: 0.84831905
00:20:07.560 --> 00:20:10.520 That basically means 000 our
NOTE Confidence: 0.84831905
00:20:10.520 --> 00:20:13.480 way through and what is.
NOTE Confidence: 0.84831905
00:20:13.480 --> 00:20:14.890 Research is probability.
NOTE Confidence: 0.84831905
00:20:14.890 --> 00:20:19.060 So that is what we found that if one
NOTE Confidence: 0.84831905
00:20:19.060 --> 00:20:22.196 reopen in 14 days after the first day,
NOTE Confidence: 0.84831905
00:20:22.200 --> 00:20:23.175 observe 0 case.
NOTE Confidence: 0.84831905
00:20:23.175 --> 00:20:26.238 So that means it can be zero and 120
NOTE Confidence: 0.84831905
00:20:26.238 --> 00:20:28.723 again in this type of situation then
NOTE Confidence: 0.84831905
00:20:28.723 --> 00:20:31.367 the researchers probability is 97% 
NOTE Confidence: 0.84831905
00:20:31.370 --> 00:20:33.848 and if one observes the zero case
NOTE Confidence: 0.84831905
00:20:33.848 --> 00:20:36.204 for 14 consecutive days and then
NOTE Confidence: 0.84831905
00:20:36.204 --> 00:20:38.244 the resurgence probability is 32%. 
NOTE Confidence: 0.84831905
00:20:38.250 --> 00:20:41.140 So what is tell us is we need to be
NOTE Confidence: 0.84831905
00:20:41.222 --> 00:20:44.360 management and don’t reopen too early.
NOTE Confidence: 0.84831905
00:20:44.360 --> 00:20:46.460 So this is happened last.
NOTE Confidence: 0.84831905
00:20:46.460 --> 00:20:48.724 May and many of the state in the
NOTE Confidence: 0.84831905
00:20:48.724 --> 00:20:50.877 South do it reopened too early.
NOTE Confidence: 0.84831905
00:20:50.880 --> 00:20:53.386 Then we saw those cases are searched
NOTE Confidence: 0.84831905
00:20:53.386 --> 00:20:56.417 in the in the summer in the South.
NOTE Confidence: 0.84831905
00:20:56.420 --> 00:20:59.138 So what’s the take home away?
NOTE Confidence: 0.84831905
00:20:59.140 --> 00:21:01.940 Take a take away message on the
NOTE Confidence: 0.84831905
00:21:01.940 --> 00:21:05.478 number 2 is to control the pandemic.
NOTE Confidence: 0.84831905
00:21:05.480 --> 00:21:07.970 A single control measures not enough
money to use multiple control measures, and including the mask, wearing social distancing and massive testing, contact tracing and also supported isolation and quarantines and also effective treatment and also the vaccine. And so the in the JAMA paper we call it multi faceted. Intervention and then later on in the summer and people give it a nice name and called the Swiss cheese model. So that is a nice name and so the challenge is we we know those 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 public house control, 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.

Under so we also.

Last spring we also developed a

website on that help estimate the

RT value as a different resolution

at the for different countries,

States and counties and so you can see that.

But we copies are key curve and

for a different.

Reach reaches and so this work

was led by Andy Xu, my student,

and she lucky there my poster.

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, An infectious component, so here is supposed. Why is the number of cases so think about the number of cases and for each day in Connecticut and then
one first need to calculate this
Lambda T and this Lambda T is called.
Basically calculates the number
of people at risk,
so that is calculated using the products
of the serial interval distribution.
Multiply the number of cases in the
previous period said like 7 days and
then the Ark is a parameter one moment.
Estimate so in the original model,
the estimate RT at each time point
that estimate a lot of parameters,
and then when building this
person model and with Lambda T as
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 accommodate on the Covid features. On the zero interval distribution and from this comma distribution using the paper in publishing in nature method and then in order to estimate $R_T$ as many values and we assume a curve and the estimate by using a spline. So there are few angles in work. And so we want to estimate $R_T$ as a function but cover it and also the traditional epidemic model.
00:24:58.010 --> 00:24:59.264 One assumed answer.

00:24:59.264 --> 00:25:01.772 Him and trade is a constant

00:25:01.772 --> 00:25:03.090 overtime and so the,

00:25:03.090 --> 00:25:05.045 but in practice the entertainment

00:25:05.045 --> 00:25:06.609 rate is not constant,

00:25:06.610 --> 00:25:08.174 especially when the number

00:25:08.174 --> 00:25:10.129 of tests are goes up.

00:25:10.130 --> 00:25:12.135 What number of positive test

00:25:12.135 --> 00:25:14.679 rate goes up and then uncertain

00:25:14.679 --> 00:25:17.427 manner it will get better and.

00:25:17.430 --> 00:25:19.974 So we want to answer payment way to

00:25:19.974 --> 00:25:22.768 be a function of the coverage and

00:25:22.768 --> 00:25:26.021 also we want to instead of fixing the

00:25:26.021 --> 00:25:28.932 reporting deal if we want to use the

00:25:28.932 --> 00:25:30.984 data to model the reporting deley

NOTE Confidence: 0.7569009
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, 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.
the state level are key value. So just give example like for California you can see that the art in California is about .67 and so does this very nice banded curve for the cases and also for that. So now let me talk about the what are the factors associated with Covid infection. So as we start from the Wuhan data, then I’ll move to the US data. So the data we estimated the attack rate on the my age. So you can see that the each of the period separately, and so you can see that for the older people that purple and yellow.
and then the tax rate was. Much higher than the younger people, and so this is a good lesson. 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 pees. 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. No, the this picture was taken from my talk. 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
help the community.

And also in the spring on the so I did something that station are supposed to do and so that we spend quite a bit time how working with the state of Massachusetts and also with abroad and so helping shifting the PPE under swap on from China.

And so I was on the state, Massachusetts. The task force in the spring and then like. One thing I was really touched last spring was.

Many, many peoples and step in to help without asking expecting any credit. So they really a wonderful experience 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 Covid-19 symptoms and behaviors and also testing a result.

And so this was in collaboration with some junk. Many of you probably know him by his work in CRISPR editing and also gene editing and CRISPR.
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 presents analysis challenge, because the people who were tested or likely to be sicker. 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 exposures and 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,
and so 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 means is household transmission is dominant is prevalent. And also we found the most important symptoms 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, taste 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 --> 00:35:34.120 smell,
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 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
overcome vaccine hesitancy.

I’m going to focus on this one.

And the second defining challenge is

the massive scalable testing and so.

PCR test yes, a gold standard,

but it is expensive and to do

the massive regular testing.

So I’m going to talk about

efficient testing strategy using

the pooled testing and also the

other strategies rapid testing.

And the third component is the

implementation and compliance of

public health control measures.

So if you look at a quick job
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 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 --> 00:37:47.912 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.
And so after we account that some people have two doses on average about 40% and with so that’s really amazing. And you have this less than 10% if on the right you can see we have a serious equity issue. And in particular, basically nobody in Africa has been vaccinated, so that’s really not good. So the another defining challenges vaccine hesitation.

So in order to achieve the vaccine induced herd immunity,
we need to overcome vaccine hesitation.

So I’m going to present the findings and from the how we feel data show McCabe is my Postal. He take a lead in this work in collaboration with many colleagues.

So here is a way, a lunch, the Maxim question in how we fill up in early December.

So we develop a partnership with Kinetica and last spring and so that’s why you can see we have more respondents and in the kinetica, and also because the countries
NOTE Confidence: 0.77242935
00:39:13.350 --> 00:39:14.990 is located in California.
NOTE Confidence: 0.77242935
00:39:14.990 --> 00:39:18.206 So we had more respondent respondent.
NOTE Confidence: 0.77242935
00:39:18.210 --> 00:39:19.132 In California,
NOTE Confidence: 0.77242935
00:39:19.132 --> 00:39:22.820 so if you look at overall vaccine hesitancy,
NOTE Confidence: 0.77242935
00:39:22.820 --> 00:39:23.742 hesitancy read,
NOTE Confidence: 0.77242935
00:39:23.742 --> 00:39:27.430 you can see like thoughts are more hesitant,
NOTE Confidence: 0.77242935
00:39:27.430 --> 00:39:30.178 and so overall the vaccine hasn’t
NOTE Confidence: 0.77242935
00:39:30.178 --> 00:39:33.422 hesitancy rate is about 1818% from the
NOTE Confidence: 0.77242935
00:39:33.422 --> 00:39:36.632 hallway field data and 82% on the.
NOTE Confidence: 0.77242935
00:39:36.632 --> 00:39:39.338 What said they are likely were
NOTE Confidence: 0.77242935
00:39:39.338 --> 00:39:42.138 more likely to take the vaccine.
NOTE Confidence: 0.77242935
00:39:42.140 --> 00:39:44.490 So if you look hard, um,
NOTE Confidence: 0.77242935
00:39:44.490 --> 00:39:47.220 vaccine hesitancy rate by race and ethnicity,
NOTE Confidence: 0.77242935
00:39:47.220 --> 00:39:50.028 then you can see that people of color
NOTE Confidence: 0.77242935
00:39:50.028 --> 00:39:53.478 are much more likely to be vaccine hesitant.
NOTE Confidence: 0.77242935
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. To overcome vaccine hesitancy, so here are the results. Who are more likely to be vaccine hesitant and so we found the younger people are more likely to
NOTE Confidence: 0.84208
00:40:34.436 --> 00:40:37.427 be a vaccine hesitant and females,
NOTE Confidence: 0.84208
00:40:37.430 --> 00:40:39.992 and also health care worker essential
NOTE Confidence: 0.84208
00:40:39.992 --> 00:40:42.959 workers and also the people of color.
NOTE Confidence: 0.84208
00:40:42.960 --> 00:40:43.770 In particular,
NOTE Confidence: 0.84208
00:40:43.770 --> 00:40:46.605 black people are 3.5 times more likely
NOTE Confidence: 0.84208
00:40:46.605 --> 00:40:49.326 to be vaccine hesitant than white,
NOTE Confidence: 0.84208
00:40:49.330 --> 00:40:51.920 and people with pre existing
NOTE Confidence: 0.84208
00:40:51.920 --> 00:40:53.992 conditions and low income.
NOTE Confidence: 0.84208
00:40:54.000 --> 00:40:57.423 And also rural areas and also the
NOTE Confidence: 0.84208
00:40:57.423 --> 00:41:01.024 thoughts and also places with high kufic
NOTE Confidence: 0.84208
00:41:01.024 --> 00:41:05.040 burden and also the people who they are.
NOTE Confidence: 0.84208
00:41:05.040 --> 00:41:08.888 So those are more likely to be vaccine
NOTE Confidence: 0.84208
00:41:08.888 --> 00:41:12.004 hesitant people who wear masks and
NOTE Confidence: 0.84208
00:41:12.004 --> 00:41:14.579 also use the protective measures.
NOTE Confidence: 0.84208
00:41:14.580 --> 00:41:19.748 They are less likely to be vaccine hesitant.
NOTE Confidence: 0.84208
Talk to us in summary.

So the vulnerable group are more likely to be vaccine hesitant, and so they include people of color health care worker, essential worker and the young people female and the regions with high kovik burdens, and also people with pre existing conditions, parents and low income. And also people not using the protective measures. So an Irish last late last year, the lunch, a community engagement alliance. And so this is, uh,
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 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.
to reach the zero kovik. So what that mean is. We need to be careful and reopen slowly when the number of cases are sufficiently small and also with the control measures are so if when we opened too early and we slipped in the control measure like what happened last summer and in the South, and is likely to see the researchers and also is important to pay attention to the long color, the long term effect especially among the young people and then also the to a build. I’ve seen uptick and it’s important 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

6 four ABCDEF and then recycle 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 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 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.

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.

the basic idea is we want to

make the spell is that possible?

So for example like here you can see
that assigned person A to pull a BE

person B person to pull CD person

3 two pull ENF person four to pull.

PNC Person 5 to pull the D&F and so on and so this. This idea is after you.

Each pool has four samples and

so you can see for the 1st pool

and the test is negative.

The second pull the test is positive.

3rd pool passes positive and so on.

Then afterwards we do the pool testing

and then we can decode to see that

which person is likely to be a case.

And here you can see that.

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 and so you can think about this as the six pools are the six. Vertex is under the edges are the people and soap example like a person. 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 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 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.
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 then also comparing the hyper design with a radius and efficient, the sensitivity is pretty similar and also when we have 384 subject per batch 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 is optimal and in terms of allocating resources and so here we plot out the X axis is the total number of sample collect each day. Suppose each day we collect 3000 samples. Suppose we only have the resources to do 12 foot tests. 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 Weaponer last December and so this is every two weeks and on Thursday from 12:00 to 1:00.

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 in Indian and Harvey Fineberg. 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 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 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?

This is 96 so 96 / 6 and then then you can see that that is. How many?

How many fewer tasks it less than 20? How many fewer tasks it less than 20?

If you do individual test that is 100.
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 you do test and then basically you need to layout the sample in a batch. If you think about it already you have to think about. Basically they have eight. 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. OK, thank you yeah.
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 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 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 compare with the individual test. The sensitivities are similar and right now I’m thinking if each individual does sensitivity they can be test by individual test one this individual mix with five other. Cure lung disease samples. Basically, the concentration is diluted,
so how does sensitivity will be keep

the same and how the next ways how
to compare if there’s one positive

case with five individual their

normal cases an A normal situation

controls and compare with all six.

There just get exposure

with low concentration.

So there will be probably have

some sensitivity issue if

pulling together compared to

individual tests, then the error for

measurement error testing error for

the two different types of tests.

How did you consider them additional
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 --> 01:01:60.869 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 positive case,

so we may have some testing error

here. Yeah yeah, so yes.

Yes, I’m sorry.

Can I just jump in for a second?

So ’cause I think the comparison as

you could compare this hyper design

to just testing everybody or the

hyper design to the traditional
pool testing approach where you just divide 100 people in each one is in a single batch and I think she hung what you’re saying and it makes sense to me intuitively is by repeating people in multiple batches were increasing the chances of having doubles and triples in the same batch. And then lowering the chance of having false negatives as opposed to the traditional design where you take the 100 people and they’re only in one match. Is that right? Yeah, that’s so that’s right. So you don’t want to put a hundred 100 people in one batch,
01:03:34.450 --> 01:03:36.290 because if you do that,
01:03:36.290 --> 01:03:39.080 then the sample need to diluted a lot and
01:03:39.080 --> 01:03:41.830 then you will sacrifice the sensitivity.
01:03:41.830 --> 01:03:44.044 So that’s why when want to
01:03:44.044 --> 01:03:45.520 do the optimal design,
01:03:45.520 --> 01:03:47.860 want to account for both the
01:03:47.860 --> 01:03:49.823 balance and also sensitivity and
01:03:49.823 --> 01:03:52.154 including both of them and so then.
01:03:52.160 --> 01:03:54.824 So that’s why when we build this when we
01:03:54.824 --> 01:03:57.330 define this efficient screening capacity,
01:03:57.330 --> 01:03:58.256 this calculation.
01:03:58.256 --> 01:03:59.645 That incorporated sensitivity
01:03:59.645 --> 01:04:01.960 in the calculation as well.
01:04:02.610 --> 01:04:03.958 I see thanks yeah.
01:04:04.700 --> 01:04:06.640 So the the pool design.
So if you look at the traditional design, so here you can see that each person, basically each person is assigned to a single pool and so this is called document design, and so this design so you can see that the six for six people. In this example, like the person, one and assigned to a person to assign to P, and so this is not efficient design. And so if we assign each person to multiple pools and after decoding that will improve the efficiency. So generally the cute does should not be too big.
So here you can secure equal to 1.
That means one person assigned to one pool to equal to 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 is assigned to pull A&B and then says that’s why there’s edge here and person to assign to C&D. So this person too. And then so on.
Thanks young, I have one.

I have one last question.

If other people do not have more question so I wonder is also related to the sensitivity.

I wonder how we considered too instead of using on the testing directly but construct some posterior for each person. Use other covariates including your past history.

Whether you have higher risk an.

I wonder if we use such personalized information combined with this testing results, can we have better sensitivity?

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.

NOTE Confidence: 0.8241894

Sand were collected and.

NOTE Confidence: 0.8241894

In the testing,

NOTE Confidence: 0.8241894

but those data cannot be used for research.

NOTE Confidence: 0.82315105

Thanks, I don’t know if the

NOTE Confidence: 0.82315105

audience have further question.

NOTE Confidence: 0.82315105

Maybe you can also email see how afterwards.

NOTE Confidence: 0.82315105

So we’re running a little bit overtime,

NOTE Confidence: 0.82315105

but it’s very. This wonderful talk.

NOTE Confidence: 0.82315105

Can we have learned so much

NOTE Confidence: 0.82315105

from Seahawks services talk?

NOTE Confidence: 0.82315105

Thank you again. Thank

NOTE Confidence: 0.82315105

you very much.