Hi everyone, welcome to the sixth seminar of our seminar series on COVID-19 organized by the Department of Biostatistics at Yale University.

I'm very pleased to have here today Nicholas Christakis.

He's a senior professor of social and medical science at Yale university.

He's very well known for his research on social networks and his recent work focuses on how essentially human biology and health affect and are affected by social interaction, social networks.

So today, he's gonna talk a little bit about the epidemiology of COVID-19.

He's gonna give us an overview and updates and then he's gonna talk about his recent paper just published in Nature on how to use mobility data and population overflow.
from Wuhan to predict the spread of the COVID-19 in all the areas of China.

And then finally he's gonna talk about the new shining app called Hunala which is gonna use network science to essentially develop network sensors to four epidemic forecasting.

So Nicholas is gonna take question anytime. So you're welcome to write questions in the chat box. And I will try to monitor it and read them to him. Or you can just unmute yourself and ask questions any time.

- Well raise your hands, I can monitor the participant list for raised hands electronically raised.

That's easy for me.

- All right.

So Nicholas, thank you for participating and why don't you take it from here.
- Thank you Laura.

Thank you so much.

I see many names that I recognize on this big panel in front of me.

I'm gonna talk without slides because I find it very stressful and weird to use slides on Zoom.

I find Zoom like probably many of you do pretty weird already.

It's so disembodied and taxing in some ways.

So I'm just gonna tell you a little bit about the epidemiology of Coronavirus as it has come to be known by many people around the world in the last few months since the epidemic started.

And some of these things may be very simple or known to you, others will not be perhaps known to you, I hope I will tell you some things you don't know.

And I'm happy to take questions at any time.
And then towards the end, I'm gonna tell you a little bit about some of the projects in my lab that have raised a number of difficult statistical questions that we are absolutely eager to collaborate with people about. And Laura has been interacting with us now for quite a number of years as I've certain others of you that I can see on this list. So we are experiencing something very unusual in our species that happens from time to time, which is the introduction of a new pathogen. We happen to be alive at a moment when a new germ is entering our species and having what is known as an ecological release. It's just like when the rats were first introduced.
to New Zealand, and they found you know, Terra Incognita,

and could just take over and do whatever they wanted.

This virus spent decades evolving in barks,

probably spent some time in pangolins

that's still being worked out.

And then in an unseen way,

almost surely in October or November in Wuhan China,

leapt into human beings and gradually spread among them

and then spread around the world.

This pathogen SARS-CoV-2

bears a strong similarity to other pathogens

that have been long circulating in bats.

And it's the seventh such Coronavirus species

that afflicts us.

There are four species of Coronavirus

that just cause the common cold,

they cause about 20 or 30% of the common cold
that people get.

0:04:07.4 --> 0:04:09.21
The other viruses that
cause the common cold
0:04:09.21 --> 0:04:11.17
are other species of viruses.
0:04:11.17 --> 0:04:14.82
And two of these Corona
viruses also came from bats.
0:04:14.82 --> 0:04:18.39
In addition, there are
two prior Coronavirus,
0:04:18.39 --> 0:04:22.09
a serious Corona viruses
that have afflicted us,
0:04:22.09 --> 0:04:26.28
what is the so called SARS-1
that was pandemic in 2003.
0:04:27.3 --> 0:04:30.09
It was a kind of limited
pandemic, which I think,
0:04:30.09 --> 0:04:33.1
on the one hand, gave
certain Asian countries
0:04:33.1 --> 0:04:37.03
a taste of what could happen
so they prepared well.
0:04:37.03 --> 0:04:38.58
But on the other hand,
0:04:38.58 --> 0:04:40.21
because the pandemic petered out,
0:04:40.21 --> 0:04:42.09
it kind of lulled the rest of the world
0:04:42.09 --> 0:04:44.21
into a false sense of security.
0:04:44.21 --> 0:04:48.58
And then the seventh, before
the current SARS-CoV-2,
0:04:48.58 --> 0:04:50.04
the seventh Corona virus
0:04:50.04 --> 0:04:52.81
is something called Middle
Eastern Respiratory Syndrome,
or MERS, which is a virus that has a R naught
which we'll talk about in a moment
of less than one we think.
Each infection yields about .9 new infections.
So that epidemic self extinguishes,
which is one reason
that MERS has not become as serious as SARS-2 has become.
Anyway this SARS-CoV-2 leapt into humans sometime in November, started causing cases in December.
And by the middle of January,
the Chinese knew that it was extremely serious.
I was contacted by some colleagues in China and Hong Kong on January 23 or 24th,
about the possibility of collaborating to do some work.
We had been working for a long time,
using phone data from China to look at the impact of things
like earthquakes on shaping
people's social interactions,

or the building of high speed rail lines.

So we had a well established collaboration and well established procedures for handling data.

And they contacted me and we decided to study the impact or something to do with phone data and the pandemic.

And so we began working in earnest on the 24th of January, and all the one 58,

it reminded me of when I was a graduate student,

because we worked non stop for three weeks,

it was very exciting.

And of course, because they were on the other side of the world, you know,

I would work during the day and then hand it to them,

and then they wake up and they work during their day while I slept.
And then it would come back to me.

When we submitted the paper on February the 18th, and it was ultimately published two months later in the middle of April.

That's the paper that Laura mentioned.

And in this paper, what we did is we had phone data on 11 and a half million transits through Wuhan. We could track people as they transited through Wuhan and spread out around the country.

And we had the misfortune as a species, that this epidemic left to us at a moment in time and in a place where the one of the largest I think, annual human migration takes place. During that annual Harvest Moon Festival in China, the new year festival.

So there are 3 billion translocations of people.
that take place in China in
the run up to this holiday,

which was on January
24th or 25th this year.

So the virus steps into us at a time
when people are spreading out,

and millions of Chinese
moved throughout the country,

including transiting through Wuhan.

And unbeknownst to them, they
carry the virus with them.

And what we were able to do
is using simply the movement of people,

track with phone data, the
aggregate number of people

that left Wuhan between
January 1st and January 24th,

and spread out to the other
296 prefectures of China.

By tracking the number of people who left

and carry the germ with them,

we were able to build
a model that allowed us
to predict the timing, intensity

and location of the epidemic
up through late February.
And this model we believe, and I'll return to a little bit later, this model we believe could be useful in other sorts of situations in which there is a risk source or risk sources that one is trying to assess in terms of its impact on this spreading of an epidemic, especially if there's data available. And I'll come back to that later. So of course, these people in China in Wuhan, then of course spread out throughout the world. The Chinese were criticized for closing their internal borders by January the 25th, the Chinese had imposed stay at home orders on prefectures in China, that encompassed past 930 million people. So beginning on January 25, nearly a billion people
were under some form of home isolation.

And this really got my attention,

because the Chinese had judged

that in order to combat this pathogen,

that the enemy that they were facing

and the virus required

them to basically detonate

a social nuclear weapon.

This was how strong

they rightly in my view

felt that the epidemic was.

So they close down their own country,

but they lagged a little

in closing down travel

and leaving Wuhan.

Some people have have said

that there was some conspiracy to do that.

I see no evidence of that.

I just think they were scrambling

to cope with a pandemic,

they closed internal travel

but didn't close external travel till a week or so later.
And without of course the germ you know,

spread around the world.

Although it would have
spread no matter what.

It's in the nature of these pathogens
once they take root.

There's really no stopping them
as I alluded to earlier.

So the Chinese quarantined Wuhan
and then Hubei province surrounded home
to 58 million people on January the 24th.

Now, the first paper about this pathogen,
regarding the first 41 cases
appeared in The Lancet on the same date
around January the 24th.

And that very first paper
noted the extreme likelihood
of interpersonal spread and
the severity of the infection.

So the nature of what we were confronting
was well understood by
scientists early in January.

I don't think we can
claim that we had no idea

there was interpersonal spread,
or that it was serious.

And the virus we now
know from genetic studies

arrived in Seattle already
by the middle of January.

And this is one of the
reasons that border closures

are so ineffective as other scientists

including Neil Ferguson's group,
and Mark Lipschitz's group

have also looked at

is that by the time you're
aware of what's happening,

and you try to close the
borders, it's too late.

The pathogen has spread,
you know, surreptitiously

and cross the borders.

And in fact, it had arrived in Seattle

by the middle of January,

and via Italy, in New York
City by the middle of February.

And by then, after that point,
most of the cases throughout
the rest of the United States

actually were seated from internal cases.

And eventually, community
transmission took over

at importation, whether from
abroad or from other states

became a progressively
tinier fraction of the size

of operates at any particular location.

This is, again, typical of
what happens with epidemics.

You know, some cases
move in, epidemic starts,

and then it just takes off

and it's not doesn't matter

how many more people
come in to a location.

And of course, it's spread
into other countries

around the world as well.

Now right from the beginning,

there was a lot of effort

to estimate key epidemiological parameters

about this pathogen.
And I suspect everyone in this group knows about this,

but I'll just review quickly what's known,

and then highlight one other interesting parameter

that not as many people pay attention on

but I know this group will be interested in.

So the so called R naught, the R0

is the number of new cases

in a fully susceptible non-immune normally interacting, typical population.

That's an attempt to measure something intrinsic

about the virus in a kind of typical human population

where no one is immune,

the virus is brand new to us,

people are interacting normally,

we haven't yet taken any protective action.

So this is known as the R naught.

And for this germ, for SARS-CoV-2,

it's probably around two and a half,
and it could be as high as three.

This is high, actually.

The seasonal flu has an R naught of about 1.3 to 1.6.

Chickenpox has an R naught of 3.5 to 6.

Ebola has an R naught of 1.5 to 1.9.

And of course the champion pathogen is measles,

which has an R naught of 18,

which is why vaccination rates for measles have to be so high,

because the pathogen is so infectious,

there's a relationship between the amount of the pathogen and the required vaccination rate to stop it,

before you could get herd immunity for example.

Now, the so called Re, R sub e is the effective reproductive rate.

This is the number of new cases as the epidemic proceeds,

and as immunity rises, or as people take action.
And this number can fall and change.

So for example, if we all of a sudden became hermits,

nobody interacted with anyone else,

the Re would fall below one.

This is actually what happened to China.

They were able to track the Re and find that whereas it started at around 3 in Wuhan in January.

After their national lockdowns,

it fell to about 0.3,

each new case only created a third of a new case.

So that's, you know, when the epidemic extinguishes.

So this Re is very sensitive to the natural rise of immune people in the population,

and also the human behaviors or ultimately for lucky vaccination that we might implement.

But there's another very important parameter that I think will interest this group
and that many that perhaps not all of you have heard about,

which is the variance in the R naught, or the variance in the Re.

And there was a landmark paper that was published by Lloyd Smith and his colleagues in Nature in 2005, that quantify this using a dispersion parameter they called Kappa,

which seeks to quantify the interindividual variation in the R, in the reproductive rate of the pathogen.

So imagine a situation where, for everyone, every single person in the population, the R is two, each person infects two other people,

and another situation in which it is zero for many people, but let's say 50 for one person,

imagine the population, a small population.

The average R in these two situations could be the same.
But the ability of the epidemic to establish itself actually could be quite different, and would be much easier in the former case.

In the latter case, you have more super spreading events. There's a variance in the $R$ so you've got that right tail distribution, some situations where one person might infect 50 or 100 people, but also most of the cases are dead ends, most people infect no one.

So in a population of such people infected with such a germ, if one person leaves and goes somewhere else, most of the time they won't be able to establish an epidemic in the new location.

So the random movement of people from one population to another, from a risks source to another place.
won't be able to establish an epidemic.

So this dispersion parameter is actually quite important for what might happen in these types of a situation.

And it turns out that the dispersion parameter for SARS-2, what we're currently facing is smaller, the variance is smaller than the variance was for SARS-1.

And this actually is one of the things that are making SARS-2 worse for us.

Even though there are super spreading events now, they are fewer than they were for the previous pathogen.

And more often now, a move of a person from one place to another starts the epidemic and can result in it taking off.

Now, superspreading depends not only on the pathogen, but also of course on the host, attributes of the host that are immunity to the pathogen,
how irritable like some people,
let's say my cough more than other people,
so that might make me more likely
to be a super spreader than you.
Super-spreading events also have to do with the environment.
This is why a pact conferences of people are more likely to cause super spreading events,
then open air concerts and so forth.
So people try quickly to get a sense of the reproductive rate of this pathogen and they were successful.
There have been like dozens of studies now quantifying this and the summary statistic is around two and a half that I told you.
And also the dispersion parameter they tried to quantify.
Distinctly, people tried to quantify the case fatality rate or the infection fatality rate of this parameter.
And there's still ongoing of this pathogen

and there's still ongoing debate about the CFR

The CFR is the fraction of people who die conditional on their coming to medical attention,
or a little bit better definition, conditional on their developing symptoms,
something which is called the S-CFR,
the symptomatic case fatality ratio.
And we think this number is about between 0.5 and 1% still.
It could be as low as 0.3%.
But I doubt that it's any lower.
And notice that the case fatality rate is very sensitive to people's behavior.
You know, do people seek medical care?
You know, if they have mild symptoms from the disease,
they might never tell anybody.
Or it's sensitive to the ability of the healthcare system
to save their lives.

So this is not something that's sort of written in stone, but it's something that attempts to quantify how lethal a pathogen is it that we have on our hands.

The case fatality rate for the seasonal flu is about 0.1%.

So on average, about one out of 1000 people who get seasonal flu will die,

and SARS-2, the current pathogen we're facing is less deadly than SARS-1.

The case fatality rate for SARS-1 was about 10%.

Yeah, was about 10%.

So it's about about 10 times as deadly as the current pathogen.

And similarly the case fatality rate for the 1918 flu pandemic, which was very bad, was about 4 to 5%.

Now, one of the things that's interesting about this,
as many of you may know

is that actually a less fatal
disease is more difficult
to treat, to stop,
because when the disease kills us rapidly

like Ebola, the victim dies

before they can transmit the disease.

But if the disease's less deadly,

and the person is walking around

for a longer period of time,

while sick, they can infect more people.

So this this difference between SARS-2,

what we're facing and SARS-2 in 2003,

I should have mentioned

that the SARS-1 petered out

there were only eight and a
half thousand cases worldwide.

You know, it was a trivial pandemic

compared to what we're facing now.

So and I forgot how many
deaths but it was in,

you know, I think 500 or six 700 deaths
from the SARS-1 pandemic.

So the lower the fatality of this pandemic,

ironically makes it more dangerous,

lower the fatality on a per case basis,

because it can spread farther

and ultimately cause many deaths.

And in general, it's an evolutionary biology principle

that the pathogens don't want to kill us.

That is to say, pathogens do better

when they're not as deadly because they can spread.

And also variants of the pathogen

that don't kill us or don't kill us fast,

typically outstrip variants that do kill us fast.

So that's one of the reasons in general

we tend to see the evolution of pathogens
to be less severe as time goes by.

And I'll come back to this point as well in just a moment.

And then the infection fatality rate
as distinct from the case fatality rate,

is the fraction of people who get infected and die.

Not the ones that come to medical attention.

So we think that about 50% or develop symptoms,

we think that about 50% of people who get SARS-2 are asymptomatic.

And so this means that in this case,

because of that 50% number,

it means that the IFR is about half the CFR in this case.

So, half the people that get infected,

don't get any symptoms at all.

And so this makes the IFR lower by a factor of two than the CFR.

Now you can take these two parameters, the reproductive rate and the case fatality rate,

and you can put them on a little graph and then you can plot all of the pandemics that have occurred, let's say in the last hundred years,
this is a typical exercise
that epidemiologists engage in.

And if you do that, you
find very distressingly

that the SARS-2 pandemic
falls between the 1957
influenza A pandemic,

which was the second deadliest pandemic
we've had in the last hundred years.

And the 1918 pandemic,
which is the deadliest.

So, this is a serious pathogen SARS-CoV-2.

It's right there in between
the upper right corner is 1918,

it's not as bad as that,

but it's worse than 1957

when you look at these
two numerical parameters.

And in fact, it was clear
to many people in certainly

by February, that without action,

many people would die.

I think hundreds of thousands
of Americans would have died,

had we done nothing.
And unfortunately, I still think that hundreds of thousands will die.

We've already had 100,000 deaths, I think we're gonna very likely have at least another couple of hundred thousand deaths before the epidemic ultimately winds down in two or three years.

And this partly relates to the fact that we're gonna have more waves, which is a point I'll come back to.

The disease itself has very wide range of presentations, from asymptomatic to mild to critical and can affect many organ systems, not just the upper airway or the lungs, but also the heart and the kidneys and the intestinal system and so forth.

And the symptomatology is very protein as well.

People manifest a great variety of symptoms.
There are three clusters of symptoms,
most are respiratory cough, shortness of breath, fever.

Some are the musculoskeletal system, fatigue,
muscle pains, joint pains. And some are intera, diarrhea, vomiting,
nausea, and again, maybe a fever.

The case fatality rate for this respiratory,
for respiratory diseases in general typically varies with age.
And most most of the respiratory pandemics of the last century have had a U-shaped function.
So the very young and the very old are at the greatest risk of death.
Famously, in 1918, there was a W-shaped function.
There's some interesting theories if we have time and you're interested,
I can tell you what some other scientists have speculated as to why it was a W, the very young,
very old were killed and middle aged,

sort of working age young adults, 20s and 30s were killed.

And then finally, there’s an L-shaped or backward L-shaped curve.

So polio has a regular L-shape curve.

So polio pandemics, they kill the young and they sort of spare the old.

But Coronavirus has a backward L-shaped, it spares the young and kills the old and this is unusual,

very unusual actually for a pathogen, and the fatality rate rises

from about one out of 3000 people younger than 20.

The case fatality rate so conditional on getting sick,

one out of 3000 people will die to about one out of 100 for people

in their late 50s, early 60s to about one out of five for people who are older than 80.
So pretty sharp L-shaped curve.

And I found it very poignant, almost biblical actually and sweet that this epidemic spared the young because, you know, the young, the leading killer of young children is infectious disease, something like 60% of kids under five worldwide die of infections.

And the fact that this virus spared them was very pleasing to me.

And moving actually, and as a parent, I didn't have to worry about my college age kids or we have a new child but Eric and I do that's 10 years old, and so we didn't have to worry about him, which was helpful.

Now, another important part of the epidemiology of this condition is something known as the incubation period.
Incubation period is the time between being infected and developing symptoms.

And that is between two and 24 days, I'm sorry, two and 14 days, the incubation period varies between two and 14 days, more precise estimates of this have shown recently and people are studying this a lot, that 97.5% of people, if they're going to get symptoms, after being infected, get them by 11 and a half days.

So, people are still studying the details of this, but the gist is that early on, it was established that 95% of cases got symptoms within the first 14 days of infection. And this is the origin of the 14 day quarantine that we have all been practicing.

There's a different quantity known as the latency period. This is the time from infection to infectiousness,
how long between when you're infected
and can infect others
and very sadly for us
in this pathogen, unlike SARS-1 in 2003,
this latency period
can be a couple of days
shorter than the incubation period.

That means that people
can spread the infection
when they're asymptomatic.

And many estimates from China and Italy
and England suggest that
the majority of cases,
the bare majority maybe or
sometimes the great majority,
arise from this type of transmission.

So most people become
infected from other people
who don't have symptoms let's say.
The difference between these two
is known as the mismatch period.
Actually, in veterinary medicine,
that call it the Omega period,

which I think is kind of interesting

when you think about

the implications for us.

In some cases, the latency period

is shorter than the incubation period,

for example, like in HIV.

So people with HIV can

be infectious for years

before they have symptoms,

that makes the disease
difficult to control,

or the latency period can be equal to

or longer than the incubation

period, like smallpox.

You have to get smallpox vesicles

on your body before you
can infect other people.

So we can see who's infected.

And that makes quarantine so much easier

and so much effective.

So the fact that there is

a negative mismatch period,

that is to say that the latency is shorter
than the intubation on average.

And this condition is one of the things that makes it so nasty and difficult to treat.

In terms of transmission modes, there's a lot of ongoing research on this, it's clear that the primary mode is through respiratory droplets, people coughing, or speaking loudly or singing. There have been a number of super spreading operates associated with singing or yelling. But there's also evidence that there's airborne transmission which is small little parts of droplets come out of your mouth and fall down, which is why wearing a mask is effective. Airborne droplets can stay suspended in the air and spread farther. There is airborne transmission. But it's not so bad.
We don't think in this condition.

Although well, I won't go into it.

There's some examples.

There's also spread by fomites,

that's surfaces that we touch.

Although this is increasingly not seen

as a major vehicle for transmission,

there's also fecal transmission

although again, this is not a major explanation

for what's happening in the epidemic.

Now, how do humans respond to epidemics?

Well, the broad division is pharmaceutical interventions,

and so called non-pharmaceutical interventions.

We don't have any pharmaceutical interventions

really for this pathogen.

We have no vaccines for it,

although we're working on it.

We have no drugs, although Remdesivir
has recently been felt to have some benefit it's modest benefit and we don't have drugs, and in general viruses are very difficult to treat, antiviral medications generally are weak in their effectiveness. So, just like plague is an ancient threat to human beings, we have to respond to a familiar enemy with a familiar response, which is physical distancing. People have been physical distancing in times of plague for centuries. And unfortunately, that's what we have to do. We have to engage in a non-pharmaceutical interventions. There are two broad kinds of non-pharmaceutical interventions, individual interventions, things like hand washing, or mask wearing, or self-isolation and collective interventions
that required the action
of groups of people

or the state, border
closures, collective hygiene,

you know, cleaning the
subways for example,

testing and tracing, bans on gatherings,

school closures, and
ultimately stay at home orders.

And these two classes,

these non-pharmaceutical interventions
can be divided into
individual and collective,

but they can be divided in a different way

in what are known as
transmission reduction

and contact reduction.

So transmission reduction are things
that try to reduce the likelihood
that conditional on my interacting
with you, I give you the germ.

So wearing a mask or washing my hands
or sanitation measures
might be transmission reduction measures.

But contact reduction or condom use in the case of HIV is a transmission reduction intervention.

And contact reduction is when you try to reduce the amount of social mixing.

So gathering bands, self isolation, school closures, or in the case of HIV reduction and partner number, those are examples of contact reduction interventions.

And the point of these interventions however we taxonomise them, is to flatten the curve.

We've all heard about that now, but why are we trying to flatten the curve?

We're trying to spread out, we're trying to this wave is about to hit us with a new pathogen for which we have no immunity.

And the force, the compressive force of the wave is gonna hit us,
what we're trying to do is
deaden the wave, slow it down,

like build breakwaters offshore,

so maybe even if the same
amount of water comes ashore,

it will come ashore with lower intensity.

So that's what we're trying to do.

We're trying to flatten the curve.

And what we mean by that

is that we are going to
allow the healthcare system

and the supply chains time to work.

By flattening the curve,

maybe we can save more lives

by not overwhelming our healthcare system.

That's one bit.

The second reason we flatten the curve

is that it postpones some cases

and deaths into the future,

at which time we might have a vaccine,

that might prevent some of the deaths
or we might have better knowledge of how to treat the condition.

Again, reducing the total number of deaths.

So flattening the curve could reduce deaths in this way as well merely by the postponement function.

And finally flattening the curve may be beneficial because it postpones some of the cases to occur at a time when the pathogen might, if we're lucky, have mutated to be less deadly. Remember, we mentioned this earlier.

So if the pathogen has become less deadly, people will become infected in the future, we'll get a milder variant of the disease. But to be clear, what flattening the curve does not do is eradicate the pathogen. What we are doing is stopping transmission, not killing the germ.
The pathogen is still there, and it's going to come back.

It's coming back in Asia, it's gonna come back in the United States, there is no escaping from this.

The pathogen is now a feature of our environment with which we must cope.

Now, one of the features of this pandemic is the affliction of healthcare workers, why healthcare workers were at special risk.

This increased risk of healthcare workers has been noted since time immemorial.

Thucydides in the plague of Athens in 430 BC, talks about how doctors are dying in greater numbers and knew the reason.

It's because they're having contact with sick patients.

The same thing is happening in our society
and happened in China
and happened in Italy.

And part of the reason
they're at special risk
is that health care workers
in the course of caring for path people,
especially when they have not had
adequate personal protective equipment,
the lack of which has enraged me, in our society.
The reason is they get high viral inoculum.
So they're up close working with the patient,
the patient coughs in their face.
So they, you might get the germ from touching something
in the subway, or interacting with a colleague at work
who speaks loudly and some number of particles
leave that person's mouth
and enter your body, by the time those viruses
are able to multiply, your body might be able
to mount an immune system, immune response

and clamp down on the infection

so you don't get a serious infection.

But a healthcare worker getting a high viral load,

a large inoculum actually can't do that, is overwhelmed,

their body is overwhelmed.

And there's a very moving website that's tracking the needs of healthcare workers around the world who have died during this pandemic and it's growing every day.

Other places of outbreaks have been nursing homes, prisons, ships.

You've all heard about the cruise ships and of course the aircraft carrier and meatpacking plants,

which is a very interesting if you want, we can talk a little bit about the packing plants.

The burden of this illness falls harder on men.
Men are more likely to die to get it and to die. But equally likely to get it but they're more likely to die than women. And as is typical of infectious diseases, it's socially stratified, the poor and the marginalized and the sick are more prone to die of this condition.

Now, let's turn briefly to this issue of waves of the pandemic. Because I think it's a serious problem. Every respiratory pandemic, in the last century has had multiple waves. And these typically recur in the fall. Not always, I think, because of all the protests that we've seen and the rush to even before the protest, the rush to leave the lockdowns, I think we're gonna see an earlier wave.
And these waves typically come every year for two or three years until eventually the epidemic becomes endemic in us.

We saw waves in 2009 with a very mild H1N1 pandemic, there was a pandemic in 2009, but that pathogen was not very deadly.

It was actually less deadly than the flu.

So it circulated the whole world there were waves, we can see the waves of H1N1, but we didn't care because it didn't kill very many people.

The 1918 pandemic, the second wave famously came out of phase with the first wave.

There's some interesting theories as to why and was much four times as deadly as the first wave.
We have no way of knowing how deadly the Coronavirus second wave will be.

But I don't believe it'll be less deadly than the first wave for various reasons.

Now the reason for the occurrence of these waves, it's complicated.

It has to do with human behavior in part, which is the fact that people return to school and move indoors with the coming of the fall and it gets colder. It has perhaps to do with environmental factors to the extent that heat and humidity affect the spread or modify our body's resistance to the pathogen. And of course, the epidemic right now has gone to the southern hemisphere and is raging there. And Brazil is having for a number of reasons,
including that it didn't make any efforts
to do anything about it.

You know, many, many hundreds
of thousands of people
are going to die in Brazil.

Incidentally, I should mention.
I'll just take a small digression
that there's a lot of geographic variation
with these respiratory pandemics.
And we don't fully know the reason.
For example, in the 1957 pandemic,
there was a 30-fold variation
in the final attack rate,
the number of people that got the disease.
So Chile was really hard hit
and Egypt was spared in 1957.
We're gonna say that see the same thing
with this pandemic, some
parts of the country
will be very hard hit,
other parts of the country will not,
some countries in the world
will be hard hit, some will not.
Sometimes this will have to do with the temperature in the region, sometimes it'll have to do with what the nations did in response. But mostly, most of the variants will be chance, as far as we can tell from previous analyses of geographic variation in the pandemic. Anyway, Brazil is being hard hit at the moment. But the point I wanna make about this is that these waves illustrate the fundamental point. But this epidemic is going to become endemic among us. Either we will develop herd immunity, probably at around 50%, ultimately, of people will be required. And we can talk about, there's a subtle detail here. So, if you can compute the fraction of people that need to be infected.
before you get herd immunity naturally infected

0:39:44.72 --> 0:39:47.61 and naturally immune before you get herd immunity,

0:39:47.61 --> 0:39:48.58 with a little formula

0:39:48.58 --> 0:39:51.29 that relies on the R naught of the pathogen,

0:39:51.29 --> 0:39:53.63 as we mentioned earlier, in the case of measles,

0:39:53.63 --> 0:39:56.33 this epidemic has an R,

0:39:56.33 --> 0:39:57.97 let's say around two and a half,

0:39:57.97 --> 0:40:00.9 it means that about 60% of people need to be immune

0:40:00.9 --> 0:40:04.543 before the epidemic goes away.

0:40:06.209 --> 0:40:10.55 But actually, you can sometimes reach herd immunity

0:40:10.55 --> 0:40:12.2 at lower percentages,

0:40:12.2 --> 0:40:15.17 because of the fact that human populations are not well mixed,

0:40:15.17 --> 0:40:16.75 they have a structured, a network structure.

0:40:19.17 --> 0:40:24.15 So typically popular people are more likely to get infected

0:40:24.15 --> 0:40:26.28 and therefore more likely to get immune.

0:40:26.28 --> 0:40:27.88 And once they become immune,
they're no longer pathways for the movement,

0:40:30.3 --> 0:40:33.34 they're no longer vectors for the movement of the pathogen.

0:40:33.34 --> 0:40:35.63 So in fact, if you immunize,

0:40:35.63 --> 0:40:38.09 let's say, 30% of the most popular people

0:40:38.09 --> 0:40:40.79 in a population, you could reach herd immunity

0:40:40.79 --> 0:40:42.62 at lower percentages.

0:40:42.62 --> 0:40:43.91 So, in practice,

0:40:43.91 --> 0:40:46.36 what we typically find is that herd immunity

0:40:46.36 --> 0:40:48.74 is reached at a lower percentage.

0:40:48.74 --> 0:40:51.92 This is, you know, the pre-pharmaceutical era

0:40:51.92 --> 0:40:55.04 at a lower percentage than you would predict

0:40:55.04 --> 0:40:56.63 based on the amount of the R naught of the pathogen,

0:40:56.63 --> 0:40:59.19 for example in 1957,

0:40:59.19 --> 0:41:02.43 the epidemic maxed out at around 40%,

0:41:02.43 --> 0:41:03.487 was the final attack rate,

0:41:03.487 --> 0:41:07.16 you know from retrospective serology studies

0:41:07.16 --> 0:41:09.393 that were done after the epidemic.

0:41:12.189 --> 0:41:14.15
So, in our case with this pandemic,

either we will get herd immunity, or we will get a vaccine.

And I've now concluded that for whatever it's worth,

that it doesn't really matter which of those two we get to first because there'll be approximately at the same time,

the likelihood that we will be able to invent a good vaccine, fast enough, manufactured and distributed fast enough to outstrip the inevitable herd immunity seems low to me.

I do think we will get a vaccine eventually, but I'm no longer putting my hopes in that as an exit strategy for this pandemic.

Maybe we'll get lucky.

I hope I'm disproven, not disprove it, but I hope that doesn't prove to be the case.

So the attack rate if you multiply
all these quantities together

that I've been telling you,

in the end for this pandemic,

in my view will be 40 to 50%.

Maybe more probably higher if we overshoot,

which is another thing, you know,

the epidemic rages onward before we have a time to actually catch up with it.

And this partly relates to the issue of who are, you know, like I already said, the popular people and the acquisition of immunity.

Now, where do we stand with this pandemic so far?

If you look at Cyril prevalence studies to date in Sweden, which has adopted a pretty mild approach to coping with it.

Nationwide there are about 4% of people have had the disease and are now immune, I think in Stockholm
with seven or eight or 9% in the most densely populated part of Sweden.

In New York City, it's about 21%

we know from a good study,

and in various era prevalence studies,

no one has really done a perfect study at anywhere in the United States.

We've been thinking in my lab of doing such a study

in the Greater New Haven area,

picking a random sample of New Haveners,

and then following them prospectively.

Most people have done others sub-optimal.

And I don't criticize them,

it's difficult to get a random sample of people.

But if I had to guess, in our cities,

probably we are at no more than 2, 3, 4, 5%

around the country.

So if we're gonna get to an attack rate of 40,
or 50%, we have a long way to go unfortunately. I think it's important to note that the United States response to the pandemic has been awful, has been completely incompetent frankly.

And the failures in my judgment have occurred at multiple levels of government, but certainly, at the White House, has, you know, there's been an appalling lack of coordination. I think the expertise at the CDC was there, but there were men with deep expertise at the CDC and deep expertise at the National Institute of Allergic and Infectious Diseases, but it hasn't been deployed properly. But I also think it's fair to say that many of the state governments were caught flat footed.
And let's also acknowledge that many European countries that didn't have the incompetence at the level of the White House also seem to have been caught flat footed. I don't understand why. The you know, it's not a mystery. I can reach over and grab a book on my shelf that's called National Strategy for Influenza Pandemic. Many, many experts knew what was happening in January and February. And the Chinese bought us time, you know, by locking down their nation. We had two months to look at what was happening in China and become concerned. Let me tell you briefly and then I'll shut up. What are some of the projects that are happening in my lab right now
which we'd be welcome.

Welcome collaborators.

And Laura's cooperating with us on some of these things.

We have ongoing work on using big data techniques.

Laura mentioned this paper on human movements.

Many scientists are working on this right now and there are labs around the world that are famous for this.

We're trying to contribute to that in a certain way, a tracking human movements or symptom reporting, that will using various Big Data techniques, including Twitter data that would allow us to forecast the course of the epidemic, to get ahead of it, to know where it's gonna strike based on knowing what's happening. Another similar project like that, being spearheaded by another graduate student
in my lab, Eric Feltham is looking at gatherings. For example, we were very interested in the gatherings to vote the primary elections. Did they, you know, people got together to vote at polling places, did that cause a spike? This is of course highly relevant to our national security, we need to somehow have a good vote, a fair and honest vote in November and for that, in my judgment, We need to have widespread absentee balloting to allow for this. Otherwise, if people stay away from the polls, because they're afraid of the pandemic, or if they go to the polls and then become infected, either one of those outcomes is a threat to our society in my view. But similarly, I believe that the recent protests.
that we've seen after the appalling murder
0:46:23.04 --> 0:46:24.433
that we saw in Minnesota,
0:46:26.7 --> 0:46:29.06
and the rioting that we've seen,
0:46:29.06 --> 0:46:31.66
are gonna contribute to a spike in cases
0:46:31.66 --> 0:46:34.49
and I mentioned this earlier.
0:46:34.49 --> 0:46:36.63
Finally, we had just released last week
0:46:36.63 --> 0:46:37.986
an app from my lab.
0:46:37.986 --> 0:46:41.413
That's called Hunala, hunala.yale.edu.
0:46:47.16 --> 0:46:49.84
This app relies on some old ideas of ours,
0:46:49.84 --> 0:46:51.88
involving network science
0:46:51.88 --> 0:46:53.75
that I previously discussed with Raphael
0:46:53.75 --> 0:46:56.283
and others on this call,
0:46:57.25 --> 0:46:58.71
which is that if you think about
0:46:58.71 --> 0:47:02.6
a contagion that begins
stochastically in a graph,
0:47:02.6 --> 0:47:03.65
you should have the intuition,
0:47:03.65 --> 0:47:04.553
it's obvious that the contagion
0:47:04.553 --> 0:47:07.55
as it winds its way a social contagion,
0:47:07.55 --> 0:47:09.91
winds its way a biological contagion,
0:47:09.91 --> 0:47:11.41
that winds its way through the graph
0:47:11.41 --> 0:47:13.23
is gonna reach central people
sooner than it reaches random people in the population.

So if we could identify central people, and monitor them, they would function as a kind of canary in a coal mine forecasting the state of the epidemic at some future time.

We published several papers about this 10 years ago, that showed that we could do it with each one and one, that we could track central people, and that we can monitor them.

And then we would therefore be able to use them to predict the future state of the epidemic between two and six weeks in advance.

So this is one of the things that our app is attempting to exploit. We're attempting to get people to report their symptoms and then we are attempting to monitor that or redraw the graph anonymously or privately.
We don't inform anyone for example, if you get sick, we do not inform your friends that you are sick.

But we exploit people's reports to create a kind of ways for Coronavirus.

So everyone contributes a little information saying whether or not they have symptoms.

And we can then by manipulating that information using some machine learning algorithms in partnership with a mean car buses group in the electrical engineering department here.

We can predict what your risk of getting the epidemic is in the future.

So if your friends friends friends had respiratory disease three weeks ago,

that should modify your risk.

Or if your friends had a fever a week ago,

that should modify your risk.
And so our app is attempting to collect these data and forecasts the future course of the epidemic. We expect to have, if we're lucky and if the app is widely adopted, we will have a ton of very difficult complicated data to analyze. And just because I haven't been as clear about this as I hope because I'm rushing, the app is like waves for Coronavirus. It warns you just like waves does that there's a traffic jam two miles ahead, the app can warn you that there is pathogen in your social network neighborhood. And just like in ways you might, you know, take an exit and avoid the traffic jam, with ways you might say, you know, I'm gonna stay at home now, because my risk is high.
And we give people daily assessments of their risk based on where they live.

And here we include lots of public source data and the reports of our users.

And we give them a risk of like based on where you live,

is the risk low, medium, high.

Like for example, like a fire like a forest fire prediction,

you know, based on the humidity today,

what's the risk of a forest fire?

And, we also give you a personal risk based on where you are in the network.

So traffic is bad in New York City in general,

but it's really bad on your block right now,

waves might tell you.

So our app also does that.

People have reported seeing a fire,

it's not just that it's dry and hot,

actually other users in
your area have seen a fire.

So your risk now is much higher.

So I'm gonna say one last thing,

which is that the issue is
when will this epidemic end?

And I already alluded to the fact that
its gonna end when it
becomes endemic among us,

but its also gonna end,

or epidemics have a biological
end and a social end.

And basically, the epidemic is gonna end
when we declare that its over.

But we have come to accept it.

And so I think we're
still in the early phases

of the Kübler-Ross model of grief.

You know, we have anger
and we have sadness, you know, depression.

And we have bargaining, you know,

but soon we're gonna have acceptance,

which is the only way
out from this pandemic.
Thank you.

0:51:12.1 --> 0:51:15.06
- Thank you Nicholas for this great talk.

0:51:15.06 --> 0:51:16.777
We have time for questions.

0:51:21.54 --> 0:51:23.61
- [Jose] Hello, I'm Jose.

0:51:23.61 --> 0:51:25.12
So I was curious about the app

0:51:26.676 --> 0:51:28.863
that you develop in collecting data.

0:51:29.8 --> 0:51:34.8
Are you interested in like, what network properties

0:51:35.5 --> 0:51:39.16
are you like hoping to gauge, are you looking into,

0:51:39.16 --> 0:51:43.84
like sensitivity, you know, like a core.

0:51:44.86 --> 0:51:49.86
And do you have a particular hypotheses about like,

0:51:50 --> 0:51:51.58
because like when you're in a network,

0:51:51.58 --> 0:51:54.24
you receive information,

0:51:54.24 --> 0:51:56.26
the more connected you are just like,

0:51:57.23 --> 0:52:01.16
forgot the term but I think you tend to get more information

0:52:01.16 --> 0:52:02.453
based on your connection,

0:52:03.42 --> 0:52:08.42
are you expecting that people in certain types of networks

0:52:08.44 --> 0:52:09.97
would be reporting

0:52:11.42 --> 0:52:13.44
like symptoms like earlier?

0:52:13.44 --> 0:52:15.55
Or I was just wondering like,

what are you hoping to capture using those parameters?

- So we've studied that if you want,

you can look at our 2010 paper on the H1N1.

Networks with low transitivity are at higher risk.

So individuals that have low transitivity in their neighborhood are at higher risk for getting the flu.

Unsurprisingly, people with higher degree are at higher risk,

and people that have high centrality are at higher risk.

So we've shown that before.

And eventually we hope if we reach the appropriate scale, that these parameters will also be relevant.

But it's not just the structure of the network as I alluded to earlier.

It's what's happening in the network and when.
So I don't know exactly yet.

We don't have the data yet to know what is the real signal.

is it COVID three weeks ago in your third degree alters,

is it COVID 10 days ago or 12 days ago in your alters?

You know, we don't know all these details yet.

If we get enough data, we will know the answer.

But we do know from principles that we published before

and that are mathematically predicted

what should matter.

Excuse me, I've been dying to do a K cornice forecasting.

And I've been talking to Dan Spielman,

I actually had to (mumbles) about this for years.

We have another project in Honduras

that Laura's involved with

where we're gonna look at K cornice,

I don't know whether we'll be able to do with this app.

It depends on how much use we get.
Hi, my name's Sam Burma, graduate in genetics. I'm interested in understanding a little bit more about the dispersion you're talking about with the equivalent reproductive, the effective reproductive rate of the virus, and I think I made it just missed a little bit there. But can you just go back and explain a little bit more about how this relates to the fact that there's a lower dispersion rate actually is worse?

It's a wonderful, beautiful paper by Lloyd Smith at all in Nature 2005 which is it's just, it's like one of those papers, you read it, and you get the point that you read and you get
"There's more subtlety here," and then you read it again,

"There's more subtlety here."

It's a wonderful paper.

The gist is, if you think about it,

and also just incidentally, as an intellectual point,

there's something happening in the sciences.

This is an old issue in the sciences between lumpers and splitters.

No Darwin talks about this, lumpers and splitters.

There are scientists who are concerned about getting the sense of a thing and like the average and they're scientists who are interested in variants like what's, you know, what are exceptions?

How do things spread out?

And so much of statistics and social sciences in the last 50 years has been focused
on measures of central tendency, why?

Because we invented regression models and statistical tools that were easier for us to say that, whereas variance is also so important.

And so there's a lot across the social scientists and people that are becoming much more interested in variants and in variation.

And so this is another example of that, where for a long time, people were, you know, trying to estimate the $R_0$.

And Lloyd Smith comes along at all, and says, "Wait a minute, the variance is also important."

Why does it matter?

So if everyone in the population has an $R_0$ of two,

then every single time one person goes from one place to another, then we'll restart the epidemic.
But if there's variation and some people
or most people have an \( R_0 \) naught of zero,
they cannot give it to anyone.
And one out of 100 people can
give it to a lot of people,
the epidemic is more likely to extinguish
'cause 99 out of 100 times when one person
in the latter example goes somewhere,
the epidemic stops.
Only if that's superspreader,
the person with a capacity
to be a superspreader
goes does it get started.
And if it's something
intrinsic to the germ,
then in the next cycle,
most of the transmissions
will also be zero.
And so this dispersion,
which they in that paper,
they quantified across pathogens
seems to fit with the
ability of the pathogen
to get instantiated.
Thank you so much.

And I also shared that paper in the chat for him.

Yeah, and also, just so you know,
it's estimated that with with SARS-2 what we're currently facing,
is you need four importations to get one transmission,
one community transmission.
So in fact, in Seattle patient zero,
did not infect anybody else.
The first case that arrived in the middle of January,
when they did contact tracing elaborately,
and now genetic studies,
he didn't infect anyone else.
It was a dead end.
You needed subsequent importations
before the epidemic took root in Seattle.
Same in China by the way.
Other questions?
It's very weird Zoom, 'cause I can't see you.

0:57:09.14 --> 0:57:10.52 I don't know if I'm boring you,

0:57:10.52 --> 0:57:13.76 I have no way of judging how I'm coming across.

0:57:13.76 --> 0:57:16.7 You know, I could be coming across as very aggressive,

0:57:16.7 --> 0:57:18.35 which of course not my intention.

0:57:21.592 --> 0:57:22.63 - I mean, if no one else has questions.

0:57:22.63 --> 0:57:24.51 I was also wondering what are the factors that influence our ability to detect the difference

0:57:24.51 --> 0:57:27.03 between the latency and the infectivity period

0:57:27.03 --> 0:57:29.12 'cause I remember at the early days in the US,

0:57:29.12 --> 0:57:32.72 CDC was still saying they didn't think

0:57:32.72 --> 0:57:34.3 that asymptomatic spread was a major factor for COVID-19.

0:57:34.3 --> 0:57:38.9 But now it seems like...

0:57:42.36 --> 0:57:44.05 - I haven't dug deep in what the CDC was saying but they knew

0:57:44.05 --> 0:57:46.74 there was a symptomatic transmission

0:57:46.74 --> 0:57:48.15 certainly by the end of January.

0:57:51.117 --> 0:57:53.59 There's some CDC announcements
that I haven't traced it all the way back

but for sure, by the end of January,

they were already saying this

and I think earlier in January too.

There was some confusion in December still,

but certainly by January, people knew.

The genetics of human susceptibility are very interesting, by the way.

I think we're gonna find that a small part of the geographic variation will relate to the genetics of the pathogen.

Right now, there's no evidence that yet that some strains of the pathogen are more deadly or more infectious than other streams,

a lot of interest in this topic right now.

We probably will find some of that.

And there's also some small evidence that human genes, that some people may be more of you
than others, because of, you know,

their various variants
still to be described.

So I think that's gonna be
and I think that'll be a small part
of explaining that geographic
variation, not a big part.

- Nicholas, Luke here in
the chat box has a question.

He's interested in your thoughts
about transmission settings
like nursing homes, meatpacking.

Is it network structures susceptibility
excetera driving transmission?

- Well, I think the meatpacking,
the story for the meat
packers was aerosolization
of the pathogen in a cold environment
and high density and I
have a long Twitter thread
on this if anyone is
interested on why meatpacking
and it's a worldwide phenomenon,
it's not just a phenomenon
in the United States.
So I think the explanation by the Secretary of Health, the Secretary of Health in our country, that it had to do with the living arrangements of the immigrants working in these factories, is not correct. There are other factories with other industries with similar immigrant populations and similar living conditions and they didn't have the operates. I think it's to do with the temperature in the environment where these people work. It's refrigerated, the very tight packing of the workers and up aerosolization using saws and other equipment that create aerosols that are in the turbulent wind conditions in these factories. The nursing homes is different, I think it's a very customary
health care situation

where you have very vulnerable elderly people

and health care workers that are up close

and intimate working with people.

So once the epidemic takes root, you know,

you get a very rapid spread.

I think nursing homes are more like prisons, actually,

in terms of their epidemiology,

meatpacking is different.

You know, ships,

ships are close quarters as well.

So you have young people in ships,

you know, in the US's Theodore Roosevelt,

you know, you have young healthy sailors,

although one died already,

we should acknowledge from the disease,

but they're very tight.

People are living in bunks,

you know, one on top of each other in that situation.
- Nicholas I have a question.

So we're used to think that these social distancing measure, these lockdowns, what they're really doing is reducing the number of susceptible people in the population.

So that essentially we plug in this number, this lower number in our serial models and predict the spread.

But, I think what in addition to reducing the density of the network, what they're also doing is reshaping the network because essentially the structure of the network because people don't go to school anymore.

People don't gather in bars.

So what do you think of this? Do you think we should make this information will be useful to include
in our model is right?

- Yes I do think that.

And just if there's a little detail there,

which was seen in China
and the United States

is, ironically household transmission,

cases of household transmission

typically are more severe than cases

of community acquired transmission

because of the viral inoculum idea.

If I get the pathogen from my wife,

I'm gonna get a sicker than
if I get the same pathogen

from riding in the subway,

because I'm writing something

I might get a low viral inoculum

whereas if I get it from my wife,

I'm gonna get a serious case.

You know, I kiss my wife, for example.

So intra-household transmission
is often more severe
and more efficacious than
out-of-household transmission.
So the dynamics will change in quite complicated ways, just like like you're alluding to, and I'm sure other groups are looking at this. It's not something we're actively doing.

But again, if you are interested Laura, I'm always eager to work with you, I love working with you.

- All right after this (laughing).

Now, do you do you wanna take one more question?

- I'm happy to take one more but pick one I wanna go so, let's do one more then stop.

Anyone else have something?

I'm looking around here and all the things I need to monitor.

Alright, thank you all very much.

- Thank you.

Yeah, Luke says thank you for a very interesting point as we think about contact tracing.
Okay.

And someone else asked about school reopening.

That's a long topic to keep people at the last minute.

My wife had a nice piece in The Atlantic about this,

if you're interested,

I think schools are gonna reopen,

I think they need to reopen.

They're going to reopen only 'cause we have no choice

but it would be better from pandemic point of view

if they did not.

I think that if they're gonna reopen,

a lot of procedures are gonna have to be put in place

at Yale, at nursery schools, at elementary schools.

It'll be different from place to place.

And I think that there will be a second wave.

So I think the schools will close again,

just what I suspect is gonna
happen in October, November.

1:03:42.688 --> 1:03:44.196
Thank you all very much.

1:03:44.196 --> 1:03:45.029
- Thank you Nicholas.

1:03:45.029 --> 1:03:46.54
Thank you all for joining.

1:03:46.54 --> 1:03:47.373
Bye bye.