From a colleague asking for help with planning for the intensive care unit and floor bed capacity at the Yale New Haven Hospital Health System and Yale New Haven in particular. Margret and Sohei had previously, or around the same time, been working with the statistics policy modeling an epidemiology collective on a queuing model or discussing the parameters of the queuing model for the dynamics of Covid-19 patient flow through hospitals. So we decided to use this model setup to make a concrete software product in the form of a web application that Yale New Haven Health System and other hospital systems could use for capacity planning. We wanted to respond to their very immediate need to know how full the hospital would get if Covid patients kept coming at the rates that they were seeing and how they might expand capacity to accommodate these new patients. So we created a Slack channel, a way of communicating directly in real time with the team members, who created a GitHub repository. Within, I think, only about two hours, we had a web application written in R, using Shiny framework, where you could sort of dial in the current bed capacity at a hospital system. You could enter parameters that govern the length of stay of Covid patients and how they move through...
the hospital from the emergency department
to the floor to the ICU
and then toward discharge or possibly death.
So that product went live very, very quickly.
There are many other collaborators and contributors
to the application beyond just our group.
Our goal here was to produce
something very quickly and immediately useful.
The structure of this model is shown
in this very complicated diagram.
It’s not as complicated as it looks.
The basic idea is that patients enter through
to the emergency department.
They move to the floor then to the ICU.
There are many things that could happen
if those places are full.
Each of those parts of the hospital is treated as a queue.
That is, it’s essentially a pool of patients
who are waiting to exit.
One of the ways they can exit is to step up
from the floor to the ICU.
One of the ways they can exit is to die.
Another is to be discharged
if they are no longer acutely ill.
So sort of taking into account all of this schematic, this
stylized depiction of the way Covid patients
would flow through a hospital,
we wrote a system of ordinary differential equations,
which describe formally, the dynamics of this system.
It’s a very simple type of modeling
that is very useful when the number of patients is large
and when you want sort of aggregate dynamics over time. So we’re not modeling, it’s not an agent-based model. We’re not modeling individual patients trajectories through the hospital. Rather, this idea of patient flow through the hospital, So this model, depicted schematically here, is formalized in a system about ordinary differential equations with many parameters. Those parameters are calibrated to data that we have from the Yale New Haven Health System and to values from the literature. We wrote this web application, which is now live at the Shiny apps URL that you can see below. You can interact with it if you like. It basically allows the user to specify time horizon, how quickly or slowly they think new Covid patients will present to the emergency department and then on subsequent tabs, you can dial in the current hospital capacity at your institution. You can dial in capacity increases that you anticipate being able to implement into the future to see how dynamics would change if say, you could add 100 new ICU beds over the course of two weeks a month from now, for example. Then there are many, many input parameters. Things like the age-specific rates of death or of stepping up from the floor in the ICU, to the average length of stay in each of those compartments for patients who come to the hospital.
You can generate reports, downloadable PDF reports. We sort of envisioned this tool being responsive to the needs of hospital decision makers who wanted to be able to add this planning capability to their existing bed management software applications and then to be able to generate reports for say, supervisors and higher up decision makers that would describe the scenario that the analysts was most interested in. The reports would also describe the consequences of a capacity expansion strategy that might be implemented by the system. So I think this tool was very useful to the Yale New Haven Health System. It was publicized kind of broadly and we got some interest from hospital systems throughout the U.S. I had spoke to some of them about the ways that they were making decisions, planning capacity increases and using this application and others that are also publicly available online, to help guide their decision making. This is an open source project. You can get all of the source code for the Shiny application on our GitHub repository here, shown below. So what are the next steps for this project? Fortunately, hospitalization in Connecticut is declining. This figure that I’ve shown here is kind of compressed. It’s declining slowly.
But it has been declining for I think, more than three weeks now. Yale New Haven Health System, along with hospitals health systems throughout the state, are doing much better than they were in mid-April. They have enough bed capacity to accommodate all the Covid patients and many more who may arrive in the coming months. So this is very good news for the hospitals and for the state. It’s one of the reasons that the governor initiated the first phase of the reopening plan on May 20 this week. However, a lot of the projections and some that I’ll show in a few minutes, indicate a substantial risk of resurgence of new cases, hospitalizations and deaths following reopening the state. This resurgence is anticipated to occur in July, August, maybe September, depending on how things go with reopening. So I think that the model, the web application, and this work in general will unfortunately, become useful again and very relevant again later on in the summer if hospitalization of Covid patients increases again. So we want to maintain our capacity to continue developing this model and responding to the needs of decision makers within hospital systems. We’re taking this down time though,
to write a technical report and a lessons learned paper about the way that we interact with health systems and how we might improve the way that we do that in the future. This work is of course also gotten us very interested in the ways that hospitals manage Covid patients. We’re very interested in comparative evaluation and comparative effectiveness in the evaluation of Covid-19 medical interventions.

That’s something that Margret Erlendsdottir an MD PhD student in biostat is working on.

- Forrest? - Yes?

- Can you take a question?

- Yes, please go ahead.

- Have you only applied this to Yale New Haven?

- We have, the model itself is generic. This is a good question, but we have calibrated many of the length of stay and probability parameters based on data that we received from Yale New Haven. So in that sense, the dynamics that we present by default are specific to Yale New Haven.

- The user has the ability to change all of those parameters, so we anticipate that this could be useful for hospital systems of any size with different patient demographics, different age distributions for example.

- So we want it to be as useful as possible, but having said all this, the customer in this case, was very clearly for us, Yale New Haven and they had a very specific need and...
Did you have an ongoing reaction with the people at Yale New Haven who were using this product, whether or not it was helping them or was it accurate or did they have any complaints about it? I’m sure they did. Can you tell me about that interaction?

Sure, sure. I think that they made a few requests of us. Some of them were very qualitative. They wanted very early to be able to generate reports. A lot of the requests were for additional functionality rather than additional structure in the OD model but they really, I think many of the requests were about flexibility and granularity in the predictions. They wanted to be able to dial in the exact patient demographics and the care parameters that were actually being implemented at Yale New Haven. So we tried to give them that ability and that control. I think mostly, successfully.

We retained some of the generality of the model, while allowing users to input the parameters that they felt were right for their system. In terms of the way it was used at Yale New Haven, I think that by the time they asked us for help, many of the actual capacity expansions had already been implemented. I’m talking about taking over high school gymnasias, changing the configuration of parking lots to provide drive through testing and turning, I guess parts of the hospital into ICUs.
Many of those–You might say that they over expanded a little bit since they quickly came not needed capacity. So did you help them, saying hey, you guys don’t need to do that much? I think that based on the projections for population level incidence that they were receiving in early to mid-April, the capacity expansion was appropriate. This model here did not provide population level projections, which I’ll show in a few minutes. So we were not telling them that they had over expanded capacity. I think that at the state level, the total hospitalization in the state came very close to the preexisting capacity, as it was in early March. So I think that there was a big concern that it was unclear what the doubling rate of new cases would be. We had not yet seen some of the benefits of state lockdown and closure of schools. So the hospital systems were expanding very aggressively, I think for good reason. Okay, but they were just doing that by looking at the daily or maybe the weekly case counts right, and seeing what the doubling rate was and things like that. They were doing anything more subtle than that? That is what they were doing when they called us on. We tried to give them projections under their own
in-house assumed doubling rates.
So we were very interested in showing them when the hospital would fill up and under what circumstance and how different parts of the hospital would fill up.
Okay.
I'll let you go on.
In a few minutes I'll show state level projections that might answer some of your questions.
All right.
By the way, I'm John Hardigen, by the way.
Used to be in the statistics department.
Yes, I know, good to see you.
All right, second project.
On April 14, so just as we finished the most fundamental software development on the application that I just showed you, on April 14 we were asked to start producing projections for the governor's Reopen Connecticut Advisory Panel, which was charged with making recommendations to the state, to the government,
the Department of Public Health on how reopening should proceed and what the timeline should be and what business sectors could safely reopen at which times.
The panel consisted of public health researchers, including Albert Ko and several other people from Yale and many business leaders in Connecticut.
It was a mixed group.
The panel needed projections at that time of Covid-19
incidence, hospitalizations and deaths under future reopening scenarios, to plan testing expansion,
seroprevalence studies and most importantly, to assess the risk of a second wave of infections. So this was in mid-April,
around the time when hospitalization was peaking. Of course, nobody knew exactly at that time that the peak was occurring and there was a lot of concern
that things would continue to get much worse, in terms of hospitalization in Connecticut. The work of that committee advised the governor in his reopening strategy, which we’ve all probably heard about, if you’re following press releases from the state. The state began reopening on May 20th and there’s now, I think, although the work of the advisory panel may be wrapping up, there’s now an ongoing need for projections to inform decision making and epidemiological study design, that further informs decision making for the Connecticut response and reopening. This part that I’ll talk about now is joint work with Olga Morozova and Richard Li. So at the beginning of this project, we had to explain how data are different from model projections and what sort of...
But I think there is a recognition at that time on the part of policy makers and committee members that the policy makers have access to a real-time data stream, which is very high quality. They have access to all sorts of state dashboards describing the current state of the Connecticut pandemic. They know about hospitalization and bed capacity information from the Connecticut Hospital Association. They know about test counts and nearly real-time case counts, number of tests positive at hospitals and in the community. They know how many deaths have occurred attributable to Covid-19 or that are suspicious, that are possibly related. They might have information about excess deaths that are not attributed to Covid-19 but are above and beyond what you might normally expect in a typical year. They have access to all this information. They have access to very responsive staff and many very smart people working for the state Department of Public Health and other state agencies. So there might be a sense that policy makers have access to all the information and the most timely information they could possibly need to make good decisions for the state. We tried to argue that there was more information that might be able to use constructively to guide reopening, and that was information that was not
0:16:19.11 –> 0:16:23.93 directly derived from contemporaneous data streams,
0:16:23.93 –> 0:16:26.71 but rather these would be projections from transmission
0:16:30.44 –> 0:16:32.67 So projections here can tell us about what might
0:16:32.67 –> 0:16:34.4 happen in the future, possible hypothetical
0:16:34.4 –> 0:16:39.13 or counterfactual scenarios to be defined
0:16:39.13 –> 0:16:42.44 by the governor and the outcomes that would occur
0:16:42.44 –> 0:16:43.83 under those reopening scenarios.
0:16:43.83 –> 0:16:47.4 So I’m talking about phases, business sectors,
0:16:47.4 –> 0:16:50 reopening back-to-school, what might happen in late August,
0:16:50 –> 0:16:51.995 early September, as children go back to school
0:16:51.995 –> 0:16:56.42 or back to summer camp in June and July.
0:16:56.42 –> 0:16:59.51 What might happen under expanded testing
0:16:59.51 –> 0:17:03 and contact tracing or continue to modified
0:17:03 –> 0:17:04.37 social distancing guidelines.
0:17:04.37 –> 0:17:08.4 Things like wearing masks or keeping six feet apart
0:17:09.821 –> 0:17:11.19 and all of those things.
0:17:11.19 –> 0:17:15.03 So we tried to explain how projections from these types
0:17:15.03 –> 0:17:18.821 of models might be very different from simple plots
0:17:18.821 –> 0:17:23.316 of the data streams that policy makers have access
to.
0:17:23.316 –> 0:17:26.37 This is a figure I showed them at the very beginning.
0:17:26.37 –> 0:17:28.08 On the left, we have the number of death,
0:17:28.08 –> 0:17:32.78 I think by early May, that had accumulated in Connecticut.
0:17:32.78 –> 0:17:35.72 These are the red dots on the left hand side.
0:17:35.72 –> 0:17:38.41 On the right hand side, we have a projection of what might
0:17:38.41 –> 0:17:40.42 occur in the future on this day and I think it was
0:17:40.42 –> 0:17:41.87 first week of May.
0:17:41.87 –> 0:17:46.87 Right, and I think this may seem silly as a projection
0:17:46.91 –> 0:17:48.97 exercise or it seems silly to
0:17:52.25 –> 0:17:55.84 make a distinction between data and predictions,
0:17:55.84 –> 0:17:58.65 but it may have useful in the setting to emphasize
0:17:58.65 –> 0:18:01.72 that the real-time data that policy makers were using
0:18:01.72 –> 0:18:03.73 was just the stuff on the left
0:18:03.73 –> 0:18:07.26 and that if one believed the assumptions underlying
0:18:07.26 –> 0:18:09.373 some of these dynamic transmission models,
0:18:10.22 –> 0:18:13.14 that they could be provided with the stuff on the right,
0:18:13.14 –> 0:18:14.38 which would be a projection
0:18:16.49 –> 0:18:18.63 Here, I happen to have shown projections starting
0:18:18.63 –> 0:18:23.61 on March 1, just to emphasize sort of how the line
follows
0:18:23.61 –> 0:18:26.59 the data points in the projection.
0:18:26.59 –> 0:18:28.57 But the idea is that these projections would come
0:18:28.57 –> 0:18:32.89 with some sort of uncertainty windows or sets
0:18:33.97 –> 0:18:36.213 that would represent, in some sense,
0:18:37.53 –> 0:18:41.4 the most likely possible futures under what we know
today
0:18:41.4 –> 0:18:44.49 and what we believe may happen about the future.
0:18:44.49 –> 0:18:46.532 So the – - May I stop you for a second?
0:18:46.532 –> 0:18:47.38 - Of course. - Forrest.
0:18:50.37 –> 0:18:53.1 First of all, I think you’ll agree that the points
0:18:53.1 –> 0:18:54.95 on the line at the left are extremely highly correlated
0:18:54.95 –> 0:18:57.683 with each other, since they’re just cumulatives.
0:18:58.79 –> 0:19:00.87 And that’s not a good way to show what’s happening,
0:19:00.87 –> 0:19:02.36 is to look at cumulatives.
0:19:02.36 –> 0:19:06.122 You have to kind of guess what the derivatives are
0:19:06.122 –> 0:19:07.406 and people aren’t so good at that.
0:19:07.406 –> 0:19:10.36 You would be much better off trying to project
0:19:10.36 –> 0:19:12.54 and look at say, the weekly values.
0:19:12.54 –> 0:19:15.22 Certainly can’t look at daily values because God knows
0:19:15.22 –> 0:19:16.817 what the daily values goes from,
but you know, you see in a week they kind of catch up with the truth. So if you looked at weekly values, you would tell on what the present situation was. Surely, that’s what the hospitals need to know. They don’t need to know how many people they had a long time ago or what the total was. They want to know that the present charge is. So I would just suggest that the thing you should be working on is something closer. Can’t use daily values, it’s too small, but a weekly value and then that’s what really matters. That’s the present situation. Certainly and have access to all that information. The State Department of Public Health produces weekly smoothed and unsmoothed count. In fact, daily counts as well. They’re very volatile. They jump up– The daily counts have a huge weekly effect. You just don’t want to rely on them at all. The docs aren’t bothered to do things on the weekends is my interpretation of it. But maybe it’s someone not bothering, but whatever it is, it’s a big weekly effect. It’s something you don’t want to have. But if you take a weekly value, that’s always averaged out. I just think that projecting the future and I think you would find there’s quite a lot more error in that.
You’re getting the benefit of the fact that all this is highly correlated but if you were trying to project the future, these things would be whoa, of stuff.

Yes, totally agree.

This figure was generated in response to a very specific question, which is how many deaths will the state expect to have accumulated on a future date.

Okay.

Thanks.

Okay, so I wanted to answer this question because I hope that you’re all wondering about it.

Does the world need another Covid-19 projection model?

There are lots of them out there.

Vary in quality, some from very experienced research groups and experienced epidemiologists, some from Silicon Valley software developers who just learned about regression.

I don’t think that the world needs another Covid-19 model at the national or international level.

But I think Connecticut does for several reasons that I wanted to describe briefly here.

We wanted to develop a scenario analysis tool that was responsive to specific questions from the Connecticut leadership, who were planning to reopen the state.

We thought there were several reasons that we could add some value here, beyond what is provided by some of the more generic models that are available.
0:21:47.021 –> 0:21:50.3 for national, state and also local projections.
0:21:50.3 –> 0:21:52.87 The first thing is access to epidemiologists
0:21:52.87 –> 0:21:54.19 at the School of Public Health
0:21:54.19 –> 0:21:56.81 and in the Public Health Modeling Unit.
0:21:56.81 –> 0:21:58.55 We have pretty unique access to data
0:21:58.55 –> 0:22:00.58 from the Connecticut Hospital Association
0:22:00.58 –> 0:22:04.623 on the bed capacity and bed occupancy throughout
the state.
0:22:06.38 –> 0:22:09.85 We can use information on individual patient trajectories
0:22:09.85 –> 0:22:11.76 through the healthcare system from using data
0:22:14.35 –> 0:22:17.52 We have access to empirical epidemiological studies
0:22:17.52 –> 0:22:20.66 from Yale emerging infections program and data streams
0:22:20.66 –> 0:22:23.36 from the Department of Public Health through Yale
EIP.
0:22:24.94 –> 0:22:27.99 We have connection to the people who are running
0:22:27.99 –> 0:22:30.4 the testing and seroprevalence studies
0:22:30.4 –> 0:22:33.71 to be conducted in the future and the model projections
0:22:33.71 –> 0:22:37.66 that we produce will be very closely tied
0:22:39.21 –> 0:22:42.11 Some of them can give information that we can use
0:22:42.11 –> 0:22:43.98 for calibrating the model, and in turn,
0:22:43.98 –> 0:22:45.69 we can use model projections
0:22:47.582 –> 0:22:50.63 to provide preliminary estimates of say,
0:22:50.63 –> 0:22:55.56 cumulative incidence of Covid-19 for study planning,
0:22:55.56 –> 0:22:57.733 in order to do sample size calculations.
0:23:00.47 –> 0:23:04.19 And of course, we are hoping to be able to help
0:23:04.19 –> 0:23:06.06 with the Department of Public Health’s
0:23:07.18 –> 0:23:10.15 implementation of optimal testing and sampling strate-
gies
0:23:10.15 –> 0:23:13.037 as they look for new cases and try to control outbreaks
0:23:13.037 –> 0:23:15.633 that may occur in the future in Connecticut.
So the modeling principle here, this is an infectious disease model that I’m gonna show you. It’s not a model for hospital patient flow through hospitals. But I think in introducing this to people who have not seen these models before, the operating principle is that of mass action. I think if mathematical infectious disease epidemiology has a central dogma or a single principle that governs the structure of quantitative models for infections, it’s something like the Law of Mass Action, that in a small time interval, the number of new cases that accrue is proportional to the number of ways susceptible individuals and infectious individuals can come together. This means that new cases or incidences are driven by the product of susceptibles and infectives or the number of ways that people susceptible individual can come into contact with an infected person. This general principle is what underlies all transmission models and many transmission models are compartmentalized or they are separated in space and geography or by age group or by different risk categories, but this is the essential principle. That new cases of a certain type and a certain place arise at a rate that is proportional to the number of susceptibles and infectives.
0:24:47.25 –> 0:24:49.8 The number of ways that disease can be transmitted.
0:24:52.53 –> 0:24:55.99 So we have divided the population of Connecticut
0:24:55.99 –> 0:24:58.65 into many compartments.
0:24:58.65 –> 0:25:00.93 Those who have not had the disease,
0:25:00.93 –> 0:25:02.87 those who are susceptible,
0:25:02.87 –> 0:25:05.76 those who have been infected,
0:25:05.76 –> 0:25:08.41 they are exposed but not yet infectious.
0:25:08.41 –> 0:25:11.49 So they don’t have symptoms.
0:25:11.49 –> 0:25:14.48 Those who are infectious but remain asymptomatic,
0:25:14.48 –> 0:25:15.67 those with mild symptoms.
0:25:15.67 –> 0:25:16.503 They know they’re sick
0:25:23.871 –> 0:25:27.97 Those who have mild symptoms but are successfully isolated
0:25:27.97 –> 0:25:31.06 because they realized they have symptoms or they got
0:25:32.96 –> 0:25:35.54 a viral test that told them that they are infected.
0:25:35.54 –> 0:25:37.93 So they successful isolate themselves.
0:25:37.93 –> 0:25:41.85 Those people with severe disease who are hospitalized,
0:25:41.85 –> 0:25:44.55 those who have severe disease but remain unhospitalized
0:25:44.55 –> 0:25:46.513 because there’s no space for them.
0:25:47.49 –> 0:25:50.33 This is very important in projecting deaths in the future
0:25:50.33 –> 0:25:53.603 scenario, in which we run out of hospital capacity.
0:25:55.26 –> 0:25:57.627 Then we have severe institutionalized populations,
0:25:57.627 –> 0:25:58.89 who are not in the hospital,
0:25:58.89 –> 0:26:03.29 such as people in nursing homes, correctional institutions
0:26:03.29 –> 0:26:06.37 and other long-term care facilities.
0:26:06.37 –> 0:26:08.92 Those who have been infected but did not die
0:26:08.92 –> 0:26:12.17 and are now recovered or successfully isolated
0:26:12.17 –> 0:26:14.483 and recovering and those who have died.
0:26:15.34 –> 0:26:18.31 So the idea here is to divide up the population
of Connecticut into a number of people in each of these compartments.
The model that we put together is a variation on the susceptible exposed infected and removed model. We divide up the infectious individuals into three categories that I told you about, severe, mild and asymptomatic infections. We have two different types of patients who need hospitalization. We have unhospitalized patients. We can remove patients by isolating them and they can recover after some amount of time, if they do not die. This is the basic structure of the SEIR model. The usual model structure is just this linear part, SEI and then R. We divided up into these additional components, not because we believed that these components cover every possible scenario or every possible type of illness or state of the world or state of patients, but because this is the most parsimonious model that we can think of that captures the dynamics of infection that are most likely to lead to the outcomes that a state government cares most about. Those are state-level hospitalizations and deaths and possibly cumulative incidents. Right, so this model is not intended to capture every biological or epidemiological feature of Covid-19 transmission in Connecticut. Rather, it is the simplest model that captures the features
that policy makers care most about.
It’s also structured by geography.
We found that the...
We looked at information about travel and commuting patterns throughout the state to look at where people might be mixing, where they live, where they work,
But we found that that information did not give us much more information than simple adjacency matrix of counties in the state.
We’re well aware that many people in Connecticut work or commute or travel often to New York City area.
We’ll try to accommodate that in the model or in our interpretation of the model.
Rather, the adjacency matrix of counties in Connecticut gives us much of the information that we use for the geographically dependent nature of transmission.
Basic idea–
Rhode Island and Massachusetts aren’t doing too good either.
They’re not doing well, I agree.
To avoid turning this into a very granular or national model
we are going to treat the exogenous force of infection experience by Connecticut residents as something else.
So we sort of imagined that it is subsumed into the force of infection experience by everyone in Connecticut.
I agree, both a lot of infections and a lot of heterogeneity outside of Connecticut in bordering states.
So most of this we don’t specifically take into account. The basic idea here, I’m just showing two compartments of the ODE system, the basic idea is that in county I, the number of susceptibles or the rate of new infections is governed by the number of infectious individuals in that county and the number of infectious individuals in neighboring counties. Here in beta is the transmission rate of infection. So individuals who are susceptible transition to the exposed infectious state and then to other states down the road. But these are sort of the mass action equations for a heterogeneous population in which the force of infection is coming from outside and within individual counties. I’m not going to go into a great deal of detail about the system of ODEs that is most useful here. I’ll just say that we solve in numerically. It’s a system of 11 differential equations given the parameters, which I’m just gonna bundle into a vector theta. Let Y of T given theta, be the solution to the ODE system at time T with parameters theta. You can solve this system with pretty good accuracy using modern ODE solvers. This solution—They’re just linear equations, are they? Linear right? They’re non-linear in the right hand side is non-linear in the other model compartments. Right, that’s what mass action is.
It’s proportional to the product. So OD is proportional to the product of S and I.

So it’s– On the other hand, you agree that S doesn’t change much because unless you’ve got a very fully infected population, S doesn’t change that much.

You’ve got– S is most quickly when infections are increasing most quickly and Connecticut right now, S is still pretty large.

I think cumulative incidence is between 5% and 15%. So S has not changed.

S is 85% and is gonna change. S is still at 95% to maybe 85%, something like that.

As the pandemic progresses and into the fall, if there’s another resurgence of infections, we will expect S to change quite a lot more.

If it changes a lot, then we’ll be in herd immunity territory where depletion of susceptibles plays a prominent role in altering the dynamics of the pandemic, but we’re not there yet.

But I am right in thinking that this is linear, it’s really just a matrix problem isn’t it, that we have to solve.

If it were linear it would be a matrix problem. Yeah, okay.
Right. - Yeah.

So this system is a deterministic system. Engineers, mostly and some epidemiologists, have been thinking for a very long time about principled ways of estimating parameters for deterministic system. Unfortunately, for models of this type, which is generally the case in infectious disease epidemiology, there are some serious identifiability problems. Not all parameters can be uniquely estimated from the data or infinitely many combinations of parameters that appear to fit equally well. We only observe in this case, the hospitalization and death compartments. There’s some information from PCR testing about the prevalence of infection at different times, but because the testing strategy in Connecticut and elsewhere has varied so dramatically over the last few months, we didn’t feel like we could use any information from testing alone to inform the sizes of the currently infected compartments. So basically, we’re trying to estimate many parameters for a system with 11 components using only the time series and also using a statistical estimation scheme.
to learn about elements of theta, of the unknown parameters. I wish that I could give you a more coherent statistical inference strategy in which all of the parameters were learned from the data and I could tell you that they were being consistently estimated and as the epidemic went on, we would get more and more precise estimates of each of those parameters. Unfortunately, it’s just not true. That the model structure that we need here to be able to accommodate the structure of the pandemic is more complicated than the model structure that we could possibly identify non-parametrically or semi-parametrically or even in this parametric model. So I just wanted to give you some examples of how people do this in practice. These are not exactly endorsements of statistical frameworks. The basic idea is that given theta, we can solve the ODE system, it gives us deterministic solutions at time points where we have an observation and then calibration or statistical inference essentially amounts to minimizing a loss criteria and are comparing the observed values to model predictions. The two frameworks that are most frequently used here are imposing a normal errors or gaussian errors, or equivalently minimizing at least squares type of loss function or doing this plus on maximum likelihood.
0:35:48.88 –> 0:35:50.75 estimation for elements of theta
0:35:50.75 –> 0:35:52.573 you can identify in this way.
0:35:53.425 –> 0:35:57.73 I think in this project we used
0:35:57.73 –> 0:35:59.293 the Poisson maximum likelihood.
0:36:00.52 –> 0:36:02.937 There are many things about this, one of which is that
0:36:02.937 –> 0:36:06.09 a Poisson random variable could take values
0:36:06.09 –> 0:36:08.4 that are larger than the size of the population.
0:36:08.4 –> 0:36:10.47 In practice here, that’s not what occurs
0:36:11.79 –> 0:36:14.55 because the number of infections here is small,
0:36:14.55 –> 0:36:17.87 but this is basically a framework for doing a type
0:36:17.87 –> 0:36:20.51 of statistical inference or learning about
0:36:22.374 –> 0:36:24.54 a posterior distribution on parameters
0:36:24.54 –> 0:36:27.03 from a model, which gives deterministic predictions
0:36:27.03 –> 0:36:29.53 and which doesn’t have any inherent stochasticity.
0:36:31.2 –> 0:36:33.01 The procedure that we used here, which I’m not gonna
0:36:33.01 –> 0:36:35.61 talk about in great detail here, was developed
0:36:35.61 –> 0:36:39.02 by postdoc Olga, is a hybrid approach that fixes
0:36:39.02 –> 0:36:42.19 some parameters and imposes uncertainty distributions
0:36:42.19 –> 0:36:46 on them from our prior knowledge and the literature
0:36:46 –> 0:36:48.66 and conducts Bayesian posterior inference
0:36:48.66 –> 0:36:51.24 on known parameters and initial conditions.
0:36:51.24 –> 0:36:53.597 So we try to learn jointly about parameter–
0:36:54.62 –> 0:36:55.924 Yes go.
0:36:55.924 –> 0:36:57.73 - Forrest, there’s a question people always ask of this
0:36:57.73 –> 0:36:59.63 whenever I give a talk like this,
0:36:59.63 –> 0:37:01.88 how do you determine your prior distribution?
0:37:03.42 –> 0:37:05.84 - In this case, I would say we’re in a very good position
0:37:05.84 –> 0:37:09.443 to interpret priors as literally being prior beliefs.
0:37:10.64 –> 0:37:14.39 We have for example, point estimates and confidence
We also have parameters which are intrinsic to the model but for which we have very little information. So we assign to them, what we believe qualitatively, to be an appropriate representation of our uncertainty or ignorance about those parameters under the parametrization.

But to your question—What you believe to be true then, is that right?

Oh certainly. It is a mixture of what other people believe to be true and what we believe to be true as well. So I would take a subject of interpretation to the priors here. They are subjective in the sense that we believe these uncertainty distributions. They are quantitative in the sense in that some of them come from published studies. Okay, I'm sorry, I do just a little bit longer.

You know, I know you've got a lot of parameters in here, many of which I don't know anything about, but I suspect the very important one is parameter which says what is the ratio of new cases assuming that susceptibility isn’t changing to the infection rate, right. What’s the... That’s the number, that ratio is an important ratio. New cases against the number that are infected and that number out to extract is an important number because it changes a lot, according to the conditions.
that the government sets.
Changes all the time because you’re trying to reduce contacts and effectively reducing that contacts is to change that ratio.
I assume that that’s built into the model somehow, but I would think you probably don’t know very much about how the government’s policies and whatever are gonna change that ratio.
So if you said you know, I know it’s gonna be a month from now, I’d say no you don’t.
So how do you handle it?
We certainly do parametrize that rate, that is the transmission rate that you were talking about. It’s the parameter that multiples the product of the number of susceptibles and the number of infectious individuals. That’s called beta in the model.
Beta does change over time.
It’s parametrized as a sum of step functions.
Those step functions change in their value around when the governor closes schools, which happened,
I think on March 25th and when the governor issued the stay at home order, or sorry, a little bit earlier, maybe March 20th,
I can’t remember.
Then when the governor issued the stay at home order, the stay safe stay at home order, which I think took effect on the 23rd.
So those step functions are in the model for historical interventions that were implemented by the state.
For future interventions which are implemented by the state, we are guessing. Fortunately, we are guessing using information from the people who will actually make those decisions. So I will show how we assume that that transmission rate or contact rate might change in the future under guidelines expressed by the governor and policy makers.

Right, so in the future of course, I don’t know what going to actually occur. The best I can do is ask the people who will implement the change.

I’m sorry, this is my last remark. I won’t keep on doing this, but I would think that these rates that we’re talking about, which seems to be are really critical to what happens in the model, that you and find invasion inferency you have to give a plausible, defensible probability for them, which I would find hard to do, and I also find it hard to do because I know that those rates differ huge amount in Connecticut between the different counties, that you can just see if you look at what’s happening in different counties. Those rates are different because different amount of separation and different amount of personal contact.

I think so kind of do that on an average way of all the counties, seven or eight of them,
you’d think you at least got a vary among the counties and have some number among the counties. Then if there’s a change of policy from the governor, there’d be a change in sum or expected you need to have that built in somehow here. Certainly.

In this work, I guess in all policy-relevant work, there is a constant tension between the need for parsimony and parametrization and the need for these rich ways of accommodating heterogeneity. What we have found in this setting is that we lacked the information or data to be able to separately parametrize transmission rates at the county level but that we can capture the aggregate number of cases, hospitalizations and other relevant outcomes at the state level by averaging over them. The reason is because the counties themselves have very different incidence, which actually does explain quite a lot in the differing trajectories of case counts and hospitalizations and deaths within the counties.

Hi Forrest thank you, this is very interesting. This is Donna. I have a question. Do you have, the parameters identifiable without Bayesian priors or just from the data that we have or do you need the priors in order to estimate the parameters? A subset of parameters is uniquely identifiable
by maximum likelihood or is point identified.

But really speaking, the answer to your question is no.

There are infinitely many combinations of parameters,

which fit any given loss function criteria equally well.

So we do need parameters here.

It is unfortunate and I think--

Yeah, go ahead.

Do you know like what’s the simplest possible model

that’s just identifiable from the data

and is that model useful at all or is it so simple

that it’s not even helpful?

Two parts to that question, the simplest model

that is identifiable from the data is probably one in which

there is no heterogeneity in types of infection,

no asymptomatic infection.

We just lump all those people together

and there’s only one kind of hospitalization

a certain proportion

of people transition to hospitalization.

That model is probably, has all the parameters identified.

And no, it’s not useful.

That seems to be what we have found.

But I would say, I think there are two kinds

of usefulness, right.

One is answering the questions that policy makers have

and the other one is what Charles Manski calls credibility,

that there is a need to take into account

known heterogeneity and known mechanisms
when we construct these models.

So if I produce a useful projection that a policy maker likes but I have not separated out asymptomatic infections,

then the numbers that I’m producing may become less regarded as less credible, right.

There’s always this rhetorical function of modeling beyond the numbers that are being produced,

to being able to accommodate or capture known mechanisms by which data are generated

is one way that we can produce more believable and actionable projections, right.

So I think there’s this balance right,

between parsimony and richness and also this balance between simplicity and believability of the assumptions.

So here we tried to you know, strike that balance.

If you think we’ve done it wrong, then please let us know.

- No, I definitely don’t think you did it wrong,

but it would be interesting to see how much you lose and sort of,

sort of cross validated predictability by adding in priors, as opposed to just using the data itself in a very simple model.

- Right, so–

I don’t know if you know the answer to that or not but you should probably go on and I know other people are wanting you to go on and not spend time answering a lot of individual questions and we can always talk another time.
Okay, sounds good.
The model fits pretty well, fits observe data pretty well.
Here, I'm showing projections that start on March 1st,
rather than at the current day or any intermediate day,
just to emphasize that
model projections and uncertainty intervals here,
which are point-wise 95%, I call it–
They are not proper confidence intervals.
They’re point-wise projections from draws,
using draws of parameters and initial conditions
from the posterior distribution over those quantities.
They’re not confidence intervals in the strict sense.
But they do appear to
match observed data quite well.
So I think we’re capturing dynamics that govern
what has occurred already.
We can learn quite a lot about the transmission rate
and under historical circumstances because we know
when those circumstances changed.
So we can estimate for example,
the percent decrease in transmission in Connecticut
following closure of schools and implementation
of the stay at home order.
That is what causes actually,
this downturn in hospitalizations and flattening
of cumulative deaths in the state.
So here, just to get a little bit more concrete,
the contact intervention.
This is a function that multiples that transmission rate
parameter that we were discussing.
So in early March, schools are closed, people start staying home and so this intervention drops down. The level to which it drops is a little more, it drops more than 85%, I think, or somewhere around 85%. That is an estimated quantity. So the drops in historical contact are estimated based on the changes in hospitalizations and deaths and the implied changes in new infections. Then what happens after the dotted line, that is after May 20th, this is just a scenario in which the amount of contact between individuals increases at, I think here, monthly intervals by 10% of the suppressed latent contact. Under this historical and hypothetical future scenario, we see cumulative incidence in the upper right-hand corner, projected from March 1st onward. Hospitalizations, with the dashed line, showing expanded hospital capacity in Connecticut. We see projections of deaths under this scenario, cumulative incidence as a proportion of the population size among people who are alive. So this is what you would get if you conducted a seroprevalence study in the future. We hope this is useful for planning those types of studies, and estimates of the affective reproduction number in Connecticut over time. There are two scenarios in particular that we want to show.
0:48:53.55 –> 0:48:56.96 policy makers that correspond to slow and fast reopening.
0:48:56.96 –> 0:48:59.05 Really, this is not reopening scenarios.
0:48:59.05 –> 0:49:02.78 I’m not sure what happened with this green annotation.
0:49:02.78 –> 0:49:03.99 I don’t know if you can see it.
0:49:03.99 –> 0:49:08.96 If I did that or somebody else did, but just ignore that.
0:49:08.96 –> 0:49:10.71 I’m not sure where it came from.
0:49:10.71 –> 0:49:13.32 Under slow reopening, we imagine that people
0:49:15.764 –> 0:49:19.06 release 10% of their latent suppressed contact
0:49:19.06 –> 0:49:22.02 every month and under a scenario like this,
0:49:22.02 –> 0:49:24.17 where everybody keeps distancing and everything goes
0:49:24.17 –> 0:49:29.17 very well in the state, new infections continue their drop
0:49:29.36 –> 0:49:32.19 and rise very slowly into the late summer and fall,
0:49:32.19 –> 0:49:34.853 hospitalization stays low throughout the summer.
0:49:38.58 –> 0:49:42.15 Deaths sort of begin to plateau and do not rise above
0:49:43.165 –> 0:49:45.733 10,000 by the end of the summer.
0:49:47.57 –> 0:49:49.82 Right, so this is the scenario that the state
0:49:49.82 –> 0:49:50.98 is really hoping for.
0:49:50.98 –> 0:49:54.78 It’s a slow reopening that does not substantially increase
0:49:54.78 –> 0:49:57.39 new infections with very slow rise
0:49:59.137 –> 0:50:01.96 in new infections as the state reopens.
0:50:01.96 –> 0:50:04.672 In contrast, a more pessimistic scenario,
0:50:04.672 –> 0:50:06.75 which I think corresponds more to
0:50:11.784 –> 0:50:16.62 a fast reopening, is one in which contact increases by
0:50:16.62 –> 0:50:21.62 or 10% of suppressed contact is released every two weeks.
0:50:22.06 –> 0:50:26.19 This results in a very fast resurgence of new cases,
0:50:26.19 –> 0:50:28.95 new hospitalizations and deaths by the end of the summer.
0:50:28.95 –> 0:50:31.2 This is what the governor would like to avoid
0:50:33.06 –> 0:50:34.2 when school children are scheduled
0:50:34.2 –> 0:50:36.183 to go back to school in the fall.
0:50:40.35 –> 0:50:42.38 There is a lot of interest right now in seroprevalence
0:50:42.38 –> 0:50:44.93 because of competing claims about herd immunity
0:50:44.93 –> 0:50:47.448 and how many people have been already infected
0:50:47.448 –> 0:50:50.2 and have evidence of prior infection.
0:50:50.2 –> 0:50:52.88 Under these scenarios, we can produce projections
0:50:52.88 –> 0:50:56.33 of the proportion of people in a random sample
0:50:56.33 –> 0:51:01.06 in the state, who might have evidence of prior infection.
0:51:01.06 –> 0:51:03.66 So this is very important for designing seroprevalence
0:51:03.66 –> 0:51:07.063 studies that we can use to further calibrate these models
0:51:07.063 –> 0:51:09.803 and that can be used to guide policy.
0:51:13.61 –> 0:51:17.27 I’m going to try to finish up very quickly here.
0:51:17.27 –> 0:51:18.82 There are a couple of key messages from this work
0:51:18.82 –> 0:51:20.81 that we tried to convey to policy makers.
0:51:20.81 –> 0:51:23.4 The first is that the state is doing pretty well,
0:51:23.4 –> 0:51:26.16 in terms of suppression of contact, closure of schools
0:51:26.16 –> 0:51:28.17 and the stay at home order have effectively reduce
0:51:32.37 –> 0:51:35.5 If contact increases quickly, the state’s at serious risk
0:51:39.95 –> 0:51:42.43 Real time metrics that policy makers have access to
0:51:42.43 –> 0:51:46.47 are really not going to serve as an early warning system
0:51:46.47 –> 0:51:47.663 for that resurgence.
0:51:48.56 –> 0:51:53.24 The state probably needs to be evaluating future projections
0:51:53.24 –> 0:51:56.163 under realistic contact scenarios for the state.
0:51:57.41 –> 0:52:00.85 We still have a lot of uncertainty that we tried to capture
0:52:00.85 –> 0:52:04.49 in model projections about cumulative incidence,
0:52:04.49 –> 0:52:06.29 asymptomatic fraction,
0:52:06.29 –> 0:52:09.6 how things are going to go with children,
the effects of enhanced testing and contact tracing and how contact patterns may change following re-opening.
So we are issuing a series of reports, which you can read online and we will be updating them in real time as the summer goes on.
You can find them at this URL. You can also email me and I’ll point you to them.
These are sort of continuously updated research products and I hope that they will represent the latest information from Connecticut and our latest predictions for the state as it reopens.
Also, there’s a document here which summarizes much more detail about the transmission model that I have given here in this presentation.
I’m gonna skip over this stuff about our workflow. We can talk about it later, if anybody is interested, but this is just how we transition from regular research to doing this type of very active software development.
I will end here.
I want to thank all of the people in the group and beyond, who have been working on this tirelessly over the last couple of months.
All of the products that I’ve told you about are publicly available.
You can find the source code on Git.
You can find the web application and the reports online as well.
So I’d be happy to take any questions.
- Thanks, thanks Forrest for the last part.
I think some people have some questions using the chat box.

Ken asked, “Is the model used at currently proposing used at hospital or by your medical group?”

The ICU planning app has been used, we know, and possibly is being used at Yale New Haven Hospital.
The projections for Connecticut are not intended for use in any particular hospital systems, though I think they will be of interest.

The projections will be useful in planning capacity expansion efforts, possibly at or beyond levels that we already saw in April.

So we will be generating any information that decision makers at those hospital systems think would be useful as they plan their response.

That’s a great question.

Let me see and Sherry asked, “In the first reopening model, what amount was the reopening assumed to start in?”

Exactly on May 20th, which is when the governor began the process of reopening.

It is also true that the governor has been
giving information about potential reopening plans for a very long time and that there is some change in contact as people begin to anticipate those changes in policy. I think that if you are looking at human mobility data from cell phones and other sources, you will see that people have been moving around for a while, increasing their level of activity outside of the home, even before May 20th in Connecticut. Whether that has actually resulted in a substantial increase in transmission remains to be seen but I don’t think we should assume that just because people are moving around and possibly returning to some types of work that there will be a corresponding increase in transmission.

Daniel asks, “Is the increase in incidence starting in September a cumulative effect of prolonged increase?” Can I just ask the question directly? I’m wondering, in the parts where you’re showing the two reopening models, it looked like the curve starts to go back up around August, September in the slow one. I’m wondering if that’s because you reach a threshold above a certain percentage of contact or if it’s a cumulative effect? Like, if we were to keep contact at .2 for example,
above a threshold, is there a situation which you don’t see that tail come up again?

- Yes, great question.

If you like to think in terms of the effective reproduction number, this increase just corresponds to a time about three weeks after that number goes above one.

So there is a threshold effect and to answer your question, if contact were to remain below a level that would give that value of one, then you would not see this type of resurgence.

I think as a practical matter, it is very unlikely that the state can avoid a situation where the effective reproduction number does above one.

I think this is not the stated strategy of anyone and it’s probably not, but I think it is the realistic expectation about what will happen in reality.

The reality is that the state is going to try very hard to increase a level of contact just about to that level, where they would see some local outbreaks that can be extinguished but they will try to maximize the level of contact, meaning economic activity and social mobility.

So they’ll try to get as much economic productivity without causing resurgence or large outbreak or an overrun of hospital capacity.

- Thank you.

- Thanks.

Akil here have two questions.
So the first one is are there any assumptions of the proposed population who have Covid-19 but have not been tested? There are implicit and explicit assumptions about that proportion. I think we can produce predictions for the current prevalence and also cumulative incidence but those predictions depend quite a lot on our prior assumptions about the asymptomatic faction. We don’t have very precise information about how many or what proportion of infections are totally asymptomatic and would go undetected by the healthcare system because people don’t seek testing or seek care of any kind when they’re not feeling sick. So certainly, we can try to learn about those things. There’s some information in the available case counts and in hospitalizations and deaths about that stuff, but we still have a lot of uncertainty about current cumulative incidence. I think it’s fair to say that currently prevalence is quite low in Connecticut. Okay thanks. I guess I saw something new saying they test the people because they can test other people that have the ability and then they have some estimate of the asymptomatic case. the rate of them? - Yes, that’s true.
In some very specific settings, like institutional settings like nursing homes and correctional institutions, you can test everybody and then you can learn how many infections are asymptomatic. The question then becomes of how representative those samples are compared to the rest of the state. Is it safe to take situations where people are living in very close proximity and possibly poor health conditions and to generalize all of that information to the state? I think there is some very good anecdotal evidence from prisons, from nursing homes and also testing systematic testing of healthcare workers that we can try to take into account, but it remains unclear how generalizable that information is. For example, healthcare workers may be immunologically somewhat unlike members of the general population who are not continuously exposed to different types of illness and to coronaviruses in particular. So I would hesitate to take large screening studies of nurses for example, and apply the asymptomatic fraction or prevalence or incidence in that sample to the general population.

And the second question of Akil is can Covid-19 models from different states learn from each other? I have relay the question is because currently your model is most about stating the data.
and you can validate how good the model is. Because states, maybe they have their reopening plan at different times, can this provide useful information about how good the model is by learning from different states.

- Yes, great question. It is always true that information from other contexts can be very useful if you know what is different in those other contexts.

I would love to be able to use more granular information from neighboring states throughout the northeast to inform projections from Connecticut, 'cause as we know, Connecticut is not an island and as soon as New York opens up and people start working, everything will change quite a lot, quite quickly in Connecticut.

So I would like to share information. We have focused on Connecticut here because we have very detailed information about Connecticut but no special access in Massachusetts, Rhode Island and New York. So that's why we've done it, but I think it will become very important and I always thought it would be the job of the CDC and the US to synthesize a national and local projections to gather all the granular local information and to put it all together. That has not happened in this particular pandemic. So I think everyone else is trying to scramble to aggregate information at the right levels.
1:02:32.65 –> 1:02:35.2 to produce predictions that are actionable locally.
1:02:36.64 –> 1:02:39.36 But there’s not coordination right now
1:02:39.36 –> 1:02:42.04 between groups that are doing state-specific
1:02:42.04 –> 1:02:44.42 reopening plans, unfortunately.
1:02:44.42 –> 1:02:47.96 As for whether the differences or staggered reopening
1:02:47.96 –> 1:02:50.42 can be used as a kind of instrument to identify
1:02:50.42 –> 1:02:53.822 the causal effects of reopening, I assume that’s the
1:02:53.822 –> 1:02:57.29 subtext of the question, the answer is yes.
1:02:57.29 –> 1:02:59.07 I think people are very interested in doing that.
1:02:59.07 –> 1:03:02.91 The problem is that reopening is somewhat endogenous.
1:03:02.91 –> 1:03:06.07 The states to reopening as a function of the conditions currently in the states and also obviously,
1:03:06.07 –> 1:03:08.61 as a function of the political considerations
1:03:08.61 –> 1:03:11.58 of the leadership and of the population.
1:03:11.58 –> 1:03:15.83 Right now I don’t think it’s safe to say that reopening
1:03:15.83 –> 1:03:18.77 occurs randomly in some time interval
1:03:18.77 –> 1:03:21.83 and that we can exploit that randomness in a simple way
1:03:24.488 –> 1:03:27 Certainly, some of the states that we observe
1:03:27 –> 1:03:29 reopen quickly, take Georgia for example.
1:03:29 –> 1:03:32.963 Those states are likely to see at least local
1:03:32.963 –> 1:03:36.4 and possibly very broad resurgences and outbreaks
1:03:36.4 –> 1:03:41.45 that may result in reversion to more restrictive movement
1:03:41.45 –> 1:03:45.35 conditions in those states.
1:03:45.35 –> 1:03:47.872 So I think really, there’s this going to be a long,
1:03:47.872 –> 1:03:50.18 longitudinal sequence of treatments,
1:03:50.18 –> 1:03:52.9 meaning changes in state regulations and then outcomes,
1:03:52.9 –> 1:03:57.06 which the regulators will observe
1:03:57.06 –> 1:03:59.56 and then this kink of cat and mouse game,
where decision makers try to tamp down on local outbreaks and then respond to ones that occur in the future. So we will try to learn about the effects of all those interventions and changes in policies but I think that there is cause for some skepticism in really learning a generalizable causal effect just from the time series.

I guess one last very specific question about a talk. So Paul asked, “Have you considered how real time data metrics, such as oxygen sensors from fitness trackers could effect your predictions?” Very interested in distributed measurements at the population level that could be helpful to inform some of these things. I think that we have not yet seen widespread adoption of mobile apps for self monitoring for contact tracing. There is some adoption of thermometers and oxygen sensors but as far as I know, there are no data streams that are publicly available. This is Paul Forcher, I asked the question. There are some, there’s— I’m participating in two studies. One that’s run out of by Mike Snider, who use to be at Yale who’s head of Stanford Genomics. The other one’s institute and if you have a fitness tracker that’s tracking...
oxygen levels, there’s emerging evidence that changing oxygen levels can be predictive of Covid infection before the patients are symptomatic and there’s some… So I would, those are two studies that you could connect with and I wouldn’t be surprised at all if they would share all of their realtime data that they’re collecting with you.

Yeah, that is a great idea, thank you. With these— Mike Snider’s a former Yale person, so you already have an inroad with that guy. Yeah, thank you, that’s a great idea. Okay thanks, I guess that’s all questions for the talk. If you have any questions, I guess they can talk to you. Like the audience can talk to Forrest offline. Please feel free to email me, anybody who has questions. Some people want to hear more about the talks, like you didn’t have time to cover, that I guess the interest you can talk to Forrest offline. This talk will be recorded and will be publicly available. Also, on the previous talk are also recorded. I’ll also send out a link to everyone in the School of Public Health, so if you want you can access it. Okay thank, thanks Forrest. Thanks everyone. Thanks everyone.