Transcript

Dr. Christine Bevc:

Welcome everyone. I'm Christine Bevc, task lead for RECOVER at the Administrating Coordinating Center, and I'll be your moderator for today's webinar. I'd like to welcome everybody to today's RECOVER Research and Review, or R3, webinar. For those of you who might be joining us for the first time, the overarching goal of this webinar series is really to catalyze the formation of a scientific stakeholder community within and beyond the RECOVER Consortium, and that includes fostering a shared understanding of the state of science and providing an educational resource for both RECOVER investigators and the broader scientific community of clinicians, patients, and public stakeholders.

I'm going to start by thanking everyone who submitted questions in advance during today's webinar. As Catherine mentioned, please use the Q&A feature in your Zoom window located at the bottom of your toolbar there to submit your questions. And after the presentations, our presenters will be answering as many questions as time allows, and some of them might be responding to those questions also in real time within the Q&A.

But please note, we unfortunately will not be answering any questions about clinical care. The questions will appear in an FAQ for this webinar and that's going to be posted along with the recording on RECOVERCovid.org. So, give us a couple days and then you'll see it there.

Today's webinar is part of our ongoing series RECOVER in Action. Today's session will focus on the characterization of PASC among children and adolescents. Our presenters will be addressing the question, what is the clinical spectrum of PASC, including sub-phenotypes among children and adolescents? And they'll be sharing those results using real EHR, electronic health record, real world data and data science. If you haven't already, please remember to sign up on our website to receive future announcements and updates on the series.

Our standard disclaimer. As RECOVER continues to grow, we want to remind our audience that the information presented in the seminar is intended to stimulate collaborative dialogue amongst both the RECOVER scientific community as well as study participants and other interested parties. The information may be recently published or about to publish, and as such is potentially subject to change. In addition, none of this information should be interpreted as medical advice.

So please join me in welcoming two of our RECOVER investigators joining us. It's my pleasure to introduce Dr. Suchitra Rao and Dr. Vitaly Lorman, and they're joined by our discussant, Dr. Kelan Tantisira who will help to kick off our discussion and response portion of our webinar. Our first presenter today, Dr. Suchitra Rao. Dr. Rao is an associate professor of pediatrics at the University of Colorado and an infectious diseases physician and hospitalist with the Children's Hospital Colorado.

Dr. Rao will be presenting on the clinical features and the burden of PASC in children and adolescents. And then next we'll hear from Dr. Vitaly Lorman. Dr. Lorman will share more on how RECOVER investigators are identifying pediatric Long COVID in electronic health record data. Dr. Norman's a data scientist at Peds.net in the applied Clinical Research Center at the Children's Hospital of Philadelphia. His background is in mathematics, and he received his PhD from Johns Hopkins. At his prior position he taught and conducted research on topology as a visiting assistant professor at Swarthmore College.

And then we'll wrap things up and hear from Dr. Kelan Tantisira, who will help to lead off our discussion starting with a brief synthesis of these presentations and a short series of questions to our panelists. Dr. Tantisira is a professor of pediatrics and chief of the division of Pediatric Respiratory Medicine at the University of California San Diego and Rady's Children Hospital located in San Diego.

So following our discussion, we'll open the floor to questions from our audience and as mentioned, we're going to try and answer as many of those questions as possible for today's topics and presentations. So please drop your questions in there and I'm pleased to welcome our speakers as we turn things over to Dr. Rao.

Dr. Suchitra Rao:

Great, thank you so much Christine, and thanks so much to you and your team for inviting us to speak with you today, and thanks to everybody who's joining. We're really excited to be here and just wanted to share with you some of the work that we are doing with the pediatric side of the EHR cohorts.

So if we can go to the next slide. The work that we're going to be presenting to you today is

work that is funded by the NIH's RECOVER Initiative, so I wanted to give you a little bit of context in terms of where we fit within all of the different elements and arms of RECOVER, and then the work that we are going to be presenting and just a couple of our published studies and looking forward to any questions that might come up about how we do this work, how this works relates to pediatrics, and we'll see where our discussions take us.

So if we go to the next slide. As I'm sure you're all very well aware, the pediatric community and research community at large have really acknowledged that we need to understand a lot more about the long-term sequelae of Long COVID in children. And I feel like that a lot of this research is still in its infancy. We know that children experience Long COVID just like adults, and we know that they may experience it in different ways given that children across all the different age spans may have different manifestations. And so it's important for us to think about this population of children and make sure that we recognize that things might be different. They might present with different incidences and they may look different from how these are all presenting in adults. And there's been a wide variation in just how often we see this in kids.

You may remember some of those earlier reports were looking at incidence rates something into the 40 to 60% mark, and then more recent studies might be showing different estimates. And one of the reasons why we're seeing this wide variation is because there's such heterogeneity in how we approach this topic. There's differences with case ascertainment, with methodology of existing studies, and some of these studies will not have an appropriate comparison group. So, it's very difficult to study something where you might see manifestations in people without the infection as well.

And so in response to this need to be able to better characterize PASC in children, a group of us, data scientists, researchers, clinicians at PEDSnet, and the group of us that are working within the pediatric EHR cohort formed a network to try to help answer these questions. So, if we go to the next slide.

So I wanted to give you a little bit of background about PEDSnet. This is the group that is really leading the work on the pediatric EHR cohort side of RECOVER. PEDSnet originated from eight founding institutions and has created a longitudinal data resource dating back to 2009. And, it cuts across all sorts of different pediatric diseases and all sorts of different pediatric specialties. PEDSnet, the work that we're going to presenting incorporates the institutions that you see here, so these are all large pediatric health systems across the country. There are additional sites that are a part of PEDSnet and are also going to be onboarded shortly. PEDSnet is a research network with

G to be presenting incorporates the institutions that you see here. So these are all large pediatric health systems across the country. There are additional sites that are a part of PED and are also going to be onboarded shortly. So PEDSnet is a research network within PCORnet. And so, you heard from Dr. Tom Carton and others a number of weeks ago about some of the work that PCORnet was doing to answer questions around adult PASC, and we are essentially the pediatric branch of that looking to answer questions focused in pediatrics.

So, if we go to the next slide, I just wanted to give you a little bit of background about the data that we are using and just how we had this network come together. Essentially, we use a common data model that standardizes our data into a single language. And by being able to do that, then you're able to pull this data from across all these different networks very efficiently and cleanly and effectively to be able to answer questions at scale and at speed.

There are a lot of data verification steps to ensure that the data is of the highest quality. And some of the data that you see includes things like diagnoses, it includes medication information, it includes lab values results. It includes demographics. We have procedure information. We are pulling in social determinants of health geocode data, and then now we're really starting to get better at pulling in unstructured data including natural language processing data and texts from notes.

If we go to the next slide. PEDSnet captures data for approximately 12 million children. It represents about 14 different hospitals across 11 network partners. That includes 357 primary care practices, 73 community clinics and 41 emergency departments. And annually, it represents about 10% of the nation's children.

And if we go to the next slide, just in terms of how this all fits within RECOVER and it's pediatric focus. The RECOVER initiative as you recall is really comprised of these four cores. And a lot of the work that is being done can be broken down into their perspective collaborative cohorts. These are the ones that are enrolling patients during the observational cohort studies. And then, the other branch of RECOVER is the EHR/Health Systems Branch. That one is going to be utilizing on the pediatric side close to nine million different records and

includes cases of confirmed COVID infection, those that have Long COVID diagnosis codes, as well as control subjects who don't have evidence of infection.

If we go to the next slide. Using this very rich pediatric data source, we wanted to answer the following questions as it relates to children. So firstly, what are the features of PASC that we're seeing in children? Secondly, what is the estimated incidents or prevalence of PASC in children? And then finally, what are some of the risk factors that we're seeing in children and is there a way that we might be able to predict who might go on to have Long COVID?

So if we go to the next slide. For this project, we included individuals who were under 21 years of age who underwent a SARS CoV-2 PCR test. And we really wanted to make sure that we captured an active population that was having and seeking healthcare at one of these PEDSnet sites. And so, we really just restricted our population to those who had at least one healthcare encounter in the prior three years to really better define this active population. Then in terms of the timing of when we were evaluating some of these outcomes, we looked at the window of time in the one month to the six-month period after the cohort entry date, and that was based on the timing of that first SARS CoV-2 test. So we had a group of patients that we would call the exposure group, so evidence of a SARS CoV-2 positive test and we had an unexposed group, which was those that were testing SARS CoV-2 negative.

If we go to the next slide. And so the approach that we took here is essentially looking at different features of PASC. So, we broke them up into syndromic features, and that really just is looking at different symptoms associated with PASC, so things like fever, fatigue, shortness of breath, palpitations, et cetera. And we looked at systemic features which were really conditions that we have been seeing in the literature emerging as being associated with Long COVID. So these are things like myocarditis, myositis, asthma ... and we used a washout period, and that's just a technique that we will employ to try and just capture those incident conditions. We know that some of these features can be found in people who don't have evidence of COVID infection, so we really wanted to find out ones that were developing after someone was diagnosed with SARS CoV-2.

And then for these analyses, we also took what we called a clinician driven approach as well as a clinician agnostic approach. And so when we did our work on the clinician driven side, we had a big team of subject matter experts, sub-specialists, pediatricians and other health partners, that were subject matter experts in various components of Long COVID. So pulmonologists, infectious diseases docs, ICU docs, et cetera. So we had a clinician work group that were pulling lists of symptoms as well as conditions that have been known to be associated with Long COVID, either based on personal experience as we were getting more information accruing or based on what we found in the literature.

And so, we developed these lists of known features to be associated with PASC. We looked at these, we had a clinical adjudication, we figured out ones that we considered to be clinically meaningful and biologically plausibly related to PASC. And so, that was what we comprised as the clinician driven approach.

The clinician agnostic approach took a list of these ICD-10 codes and whatever wasn't put into that category of clinician driven then became clinician agnostic. This was an approach that we took so that we could do some discovery as well. So rather than just relying on our clinical knowledge base, we wanted to see if there were any additional signals in the EHR data that could be useful and meaningful.

So, we started out with an AHRQ list of codes. As I mentioned, we had over 66,000 codes that we started with and we reduced those down to 441 relevant clinical features. And, our evaluation period was patients who were seeking healthcare at one of these PEDSnet sites from March 1st of 2020 through til October 31st of 2021. So, when we looked at that cohort, we had over 660,000 patients who underwent a SARS CoV-2 test by PCR across all of the different sites, and who had at least one prior visit in that seven days to three years prior to their test date. And then of those patients, a little under 59,000 of them were positive for SARS CoV-2, so that was a positivity rate of about 9%.

And so if we go to the next slide, what I'm going to show you now are just some of the results. So the next two slides are going to show you just a table that's going to be comparing those who are SARS CoV-2 positive with those that are SARS CoV-2 negative. If we look at this table here, patients who tested positive were more likely to be of older age, to be of Black race and to be of Hispanic ethnicity.

And if we go to the next slide. And, on this table, this is looking at some of the underlying

medical or chronic conditions, the time period of cohort entry and then the test location. And here, the majority of patients were tested in an outpatient setting or in an ambulatory setting. And then, a higher proportion of those inpatients who were tested were test negative.

And if we go to the next slide here. So the way that we wanted to explore different associations with Long COVID is we wanted to look at how often something occurred in the population of patients that were test positive compared to, or relative to, those who were tested negative. And the way that we did this was to use adjusted hazard ratios. So what you're going to see over the next couple of slides are forest plots that show you the adjusted hazard ratios that are plotted.

And so, the further you are to the right, the stronger the association. And then if you see something that is overlapping with one on that line, it overlaps on that line, then that may not be considered significant. And so again, we are looking at that ratio then of testing positive, having a feature amongst the positive group compared with those who tested negative. And we used an adjusted hazard ratio model and we adjusted for things like site, for age group, for sex, for race and ethnicity as well as the time period of the cohort entrance, as well as the location of testing. And so what we're just going to show you on these slides are those that are clinically relevant and statistically significant based on our findings.

So in terms of the symptoms associated with PASC, in order again from the strongest association down, we identified the following features. We found changes in smell and taste, loss of smell, alopecia or hair loss, chest pain, abnormal liver enzymes. We found generalized pain conditions as well as anxiety symptoms. There were some signals for skin rashes, for persistent fever and chills, as well as cardio-respiratory symptoms.

And then if we go to the next slide. So this was the forest plot looking at various conditions associated with SARS CoV-2 infection. And so in terms of some of our systemic features, we identified myocarditis to be a very strong signal, acute respiratory distress as well as myositis or muscle inflammation, visits for mental health treatment. We found other ill-defined heart disease thrombophlebitis or thromboembolism was strongly featured as well. And we found some associations with some respiratory and upper respiratory tract infections as well.

So, if we go to the next slide. So now we know that estimating the incidence of PASC can be challenging, and I mentioned very heterogeneous across different studies so in our attempt to try and get a sense of the incidence of PASC in this cohort, we looked at the incidence proportion difference. So, the way that we did this is we looked at the incidents for any clinically predicted or empirically supported symptom or condition or medication related PASC feature between those children in the SARS CoV-2 PCR positive group and then compared that with those who are in the test negative group.

And so, as you can see from this table here, the incidents in the test positive group was close to 42% compared to around 38% in the test negative group. So that gave us an incidence difference of 3.7%, so close to 4%.

And then if we go to the next slide, we also evaluated risk factors for PASC, and in our adjusted hazard ratio models we found some of the strongest associations for PASC to be in those who were admitted to the ICU for their acute illness. We found a strong association for children under the age of five in terms of PASC that were presenting to medical care as well as those with a chronic complex medical condition.

Now, if we go to the next slide. Now it is worth highlighting some of the limitations of using HER data for this type of work. So we looked at symptoms and signs and other diagnoses that were significant enough to prompt health service use. And so, we may not be able to capture all of the different manifestations of Long COVID in children. The burden of PASC may be underestimated from EHR data sources, and one of the reasons for this is the issue of misclassification. So what can happen is we're really just able to see those tests that occur at the health systems themselves, and so if something was taking place at a different health system with respects to testing or if all of the testing was done with at-home testing, then there may have been some misclassification and that might explain why we saw a low incidence proportion difference in this cohort.

And it's also worth pointing out that this work was done before home testing was widespread. So, that is probably a little bit less of an issue with our time period of our data. And the other important point here is that some findings may reflect differences in care seeking behavior as well as access to care.

So if we go to the next slide, so just to summarize this work again, in this study that we published evaluating the PASC features in close to 660,000 children undergoing testing for SARS CoV-2, we found that there were features that overlap with some of what we're seeing in the adult populations, but there are other

features that are maybe less well reported in adults that we're seeing in kids. And so some of the more common symptoms and common conditions are shown for you here. And this really is one of the things that highlights why having a control group is really important when you're doing these types of studies.

And we also looked at some risk factors for PASC that was resulting in health-seeking behavior. And so, we found associations like having underlying medical complex conditions, being in the ICU for the acute illness, as well as those who were under the age of five years of age.

So if we go to the next slide, so I wanted to conclude there, but then I really wanted to make sure to acknowledge all of the folks that were engaged in this work that were representing the PEDSnet sites and our subject matter experts. So, we had representatives from each of our PEDSnet sites and I also wanted to thank the team at the CHOP Data Coordinating Center, including Vitaly Lorman who's here today. Chris Forrest, Charlie Bailey, Hanieh Razzaghi, Ryan Webb and Kimberley Dickinson for all of their work on the analysis and data science. So with that, I'll pass it over to our next speaker and looking forward to discussions afterwards.

Christine Bevc:

Great, thanks Dr. Rao. So next we'll move to our presentation by Dr. Vitaly Lorman. As a reminder, if you have a question for any of our panelists, please drop them into the Q&A feature located at the bottom of your Zoom window. It's great to see those questions coming in and we're looking forward to having the opportunity to respond to some of those. Thanks. Dr. Vitaly?

Dr. Vitaly Lorman:

Thank you so much for the invitation to speak here. I'm very excited to tell you some more about work that's been going on in PEDSnet and as part of RECOVER on pediatric Long COVID and looking forward to the subsequent discussion.

So I'm going to talk a little bit about how we've been working on identifying pediatric Long COVID in EHR data, but first I want to say a little bit about why we think this question is important. I think there are really two answers. One is that identifying Long COVID in EHR data goes hand-in-hand with finding out what Long COVID is, how it presents in children, what are the associated symptoms, conditions, clinical trajectories. And additionally, being able to provide some answer to this question is necessary if you want to conduct other studies about Long COVID. For example, if you want to identify Long COVID subtypes, look at the effectiveness of various treatments, look at how Long COVID has changed over time and other studies that involve a definition of being able to identify patients who are likely to have had Long COVID.

Next slide, please. So the first approach you might take if you want to identify a cohort of patients who were likely to have had Long COVID in EHR data is you might want to see who was actually diagnosed with Long COVID. And, there are actually three relevant diagnosis codes in the ICD-10 vocabulary here. The first one is U09.9, which you may have heard about if you attended the corresponding R3 seminar for the adult EHR group. So that's for a post-COVID-19 condition unspecified, and that code was introduced in fall of 2021, so about a year and a half into the pandemic. Additionally, there's a diagnosis code for something called multisystem inflammatory syndrome in children, which is a condition that drew a lot of attention, especially earlier on in the pandemic. It tends to be very high in intensity maybe with treatment shorter in duration.

And the CDC characterizes it as inflammation following COVID-19 and roughly manifesting in at least two body systems. And because MISC is so specifically defined, we often tend to separate it in analysis. So the study I'll be talking about today, we focused mainly on non-MISC presentations of Long COVID, although then in other studies that I'll talk a little bit about at the end, we treat them all together.

And then finally there's another diagnosis code that is not specific to COVID. There's a B94.8, sequelae of other specified infectious and parasitic diseases code which predates the pandemic and it was recommended for use prior to the U09.9 code being introduced. And so, you might start by looking at patients who are diagnosed with these codes, but there are some limitations in just using these diagnosis codes which have to do with how and when these codes tend to be used.

So for example, these codes just didn't exist earlier in the pandemic, so you're unlikely to be able to find patients who had Long COVID earlier in the pandemic using these codes. And then furthermore, we're limited by the developing clinical understanding of what Long COVID is and how it's evolved over time. And so, there might be specific settings such as Long COVID clinics where these diagnosis codes tend to get used more, and maybe specific providers where they get used more and other settings where they get used less. And so because of

that, there's good reason to think that just looking at these diagnosis codes alone is likely to significantly under ascertain who has Long COVID in EHRs.

And so, the next thing you might try to do is move beyond that by identifying who has Long COVID using other data elements in the EHR. Next slide please.

And so the question you then might want to ask is what symptoms and conditions are more common in patients who have Long COVID, which I'll also refer to as PASC, Post-acute Sequelae of COVID-19. So what symptoms and conditions are more common in patients with PASC? And so to ask what is more common, you have to make a comparison to another group of patients, and there are three relevant comparisons here. You can look at patients who are diagnosed with PASC and compare them to patients who have COVID but not PASC. You could compare them to patients who didn't have COVID, and in fact had a negative SARS CoV-2 test result. And finally, you can compare not just patients who have been diagnosed with PASC but all COVID positive patients, patients with COVID, compare them to patients without COVID. And these are all important comparisons and there are subtle differences in them that all make them worth studying and they point to different things.

For example, in the third comparison, SARS CoV-2 positive patients, looking at that comparison has the advantage of looking at a much larger group of patients. There were a lot more patients who had COVID than those who had Long COVID. But on the other hand, patients with COVID-19 are a much more heterogeneous population for the reason that in the EHR, someone may have gotten a SARS CoV-2 test, not because they came in for that but incidental to something else that they were seeking care for. And so it's going to be a much more varied population.

And so in this study, our main analysis focused on the first two comparisons and then we did the third one as a sensitivity analysis, and then we tried to synthesize and investigate all of those. Next slide please.

Okay, so just to draw a little bit more attention to this idea that patients who are diagnosed with PASC might not be representative of all patients who may have had PASC or Long COVID. So, this shows in our study patients who were diagnosed with PASC compared with patients who just had COVID. And what we see looking at this table briefly is that patients diagnosed with PASC tend to be generally older than patients with COVID, so they're more likely to be in the age 12 to 15 or age 16 to 20, in those age groups. They're a little more likely to be female. They're more likely to be non-Hispanic White and in terms of presence of existing chronic conditions, we see that patients diagnosed with Long COVID are more likely to have both chronic conditions and more likely to have complex chronic conditions where we categorize this using something called the Pediatric Medical Complexity Algorithm.

And just to give an example, asthma is something which would be considered a non-complex chronic condition, and complex chronic conditions are ones that are progressive or malignant or affect multiple body systems in this categorization. So we see that patients with who are diagnosed with Long COVID do tend to differ substantially in several respects from patients who have COVID. And there are several hypotheses for this. For example, it might be that in younger patients there might be certain symptoms, for example, persistent cough that you would be more likely to attribute to early childhood illness, whereas in older patients you might be more likely to attribute them to COVID. And they might also have a lot to do, as I mentioned before, with who has access to Long COVID clinics, access to care, those sorts of things.

And so again, this is a case for moving beyond just the specific diagnosis codes for Long COVID and trying to come up with a more broad definition of what Long COVID is to try to capture a broader cohort of patients in electronic health records.

Next slide please. So I'll start now to talk more specifically about this study. In EHR data, which is secondary use data, whenever you make a comparison between two groups of patients, there's always the chance that there are incidental differences in the two groups. For example, having to do with who came in to have a COVID test recorded in the EHR or who happened to have a PASC diagnosis recorded in the EHR that might not actually have anything to do with the comparison you're trying to make. They might not have anything to do necessarily with PASC or COVID itself, but instead other differences in healthcare utilization. So because you don't want those to give you false signals in your results for what you're looking for, sometimes ... So in this study we used a method called matching, which the idea behind it is to try to control for those differences, to try to emulate a randomized control trial. Next slide please.

And so, once we had our cohorts defined, so we're comparing patients who were diagnosed

with non-MISC PASC and we're comparing them to both patients who had COVID-19 and who weren't diagnosed with PASC and to patients who did not have COVID-19, defined as presence of a negative SARS CoV-2 test. So we're making those two comparisons and the outcomes we're interested in ... there are many of them. We're interested in seeing exactly what symptoms and conditions are diagnosed disproportionately among patients with PASC compared to our two control groups. And so there are lots of possible diagnosis codes that can occur. As the Suchitra's slides showed, there are nearly 70,000 different ICD-10 codes, and so we're scanning the entire set of possible ICD codes, diagnosis codes, to see which ones occur disproportionately more frequently in patients diagnosed with PASC.

Now, diagnosis codes don't live in isolation. Instead they're part of a complex hierarchy. So the image in this slide shows underneath ... Within the ICD hierarchy, there's a branch that consists of codes that start with the letter R, which are symptoms and signs and abnormal findings not classified elsewhere. Underneath that there's a branch of codes that start with R00 through R09, which are symptoms and signs involving the circulatory or respiratory systems. And this chart here just shows that sub-branch of the ICD hierarchy, just to give an example.

And so, the method we use called the tree based scan statistic, it scans this entire hierarchy and it looks for both individual diagnosis codes but also clusters of diagnosis codes, so branches of this hierarchy, which is also called a tree, in which diagnoses in those clusters were disproportionately likely to occur in patients diagnosed with PASC compared to our control groups. And the advantage of this approach is that it can identify not only specific codes but also clusters of codes. So, it can identify for you the varying levels of granularity at which we see PASC manifested in EHR data. And then given those findings, you can then comb through them and figure out at different levels of granularity which codes you want to look for and maybe incorporate into your Long COVID definition that you want to apply to EHR based studies of PASC. Next slide please.

Okay, so now I'll just summarize at a high level some of our results. Because we were scanning so many codes, there are a lot of results here and you can find more detailed descriptions in our manuscript. So this slide shows systemic findings, so one's related to organ system function generally. And then the next slide shows syndromic findings. So one's related to symptoms. And so just to start with a high level summary, so a lot of what we found is consistent with other studies. At a broad level, we found diagnoses associated with PASC in cardiac, respiratory, neurologic, psychological, endocrine, gastrointestinal and musculoskeletal systems. And then just to highlight a few, we saw neurologic disturbances like headache and migraine recurring. We saw sleep disturbances like sleep apnea, which was a subject of a recently published RECOVER manuscript. In the musculoskeletal system, we saw multiple and diverse pain symptoms such as pain in joints, pain in limbs, soft tissue disorders that we think are worth further study in particular in terms of the underlying biological mechanisms.

And I didn't mention, but we were looking specifically for new diagnoses that occur during the post-acute period. So in other words, we were interested in diagnoses that occurred furtive for the first time at least in the previous 18 months ... for the first time after the patient had COVID or had PASC diagnosed. And so what that means is, for example, when we see specific asthma codes show up in our findings as statistically significant, that points to perhaps new onset asthma being associated with PASC. Of course it's also possible, and this is where things get kind of complicated, it's also possible there were patients who had asthma prior to having a PASC diagnosis, and with PASC perhaps their asthma was exacerbated to the point where it was recorded in the EHR but it wasn't previously. And so disentangling these things as we might get into later in the discussion, is one of the more complicated parts of interpreting these findings.

And finally, compared to previous work, including the study that Suchitra discussed, there was a lot of overlap, but we also found some differences that are worth noting. Abnormal liver enzymes, allergies, disorders of teeth and bronchiolitis were all reported as potential PASC associated features in the study that Suchitra led, but we did not see those as statistically significant in this study. And conversely, significant findings in our study that were not present in the study that Suchitra presented included sleep disorders, anemia, eye disorders, constipation, nutritional deficiencies, hypotension and various musculoskeletal conditions.

And what all of that means is really that it's worth looking at what conditions and symptoms are associated with PASC from a variety of perspectives and trying to synthesize those findings. So we expect to see, depending on what comparisons we make, how we define our cohorts, and during what time period of the

pandemic we look at these things, we expect to see different results and those differences themselves may be interesting and worth a further investigation.

Next slide please. So this slide shows specifically symptoms that were significantly more likely to occur in patients with PASC compared to our control cohort.

Next slide please. And then briefly, I just wanted to show a small sample of the more granular output of our analyses. So as I mentioned before, this tree based scan statistic approach, what it does is it looks for where there are signals, in other words where patients with PASC disproportionately tend to have diagnoses in this complex ICD hierarchy at different levels of granularity. And so, this shows just a small snippet of it where we see that, for example, the R00 to R09 branch symptoms and signs involving the circulatory and respiratory systems, in general, diagnoses in this branch tended to occur more often among patients with PASC.

But then it also points to specific codes within those clusters that might be especially driving it. For example, dyspnea, malaise and fatigue, et cetera. And so there's a very large table in our manuscript, and as you scan through it you can find where signals were detected at more general levels and where they were detected at more specific levels. And there's a lot there that might be worth investigating further in terms of what symptoms and conditions might be associated with PASC.

Okay, and then finally, the other reason that these more granular findings are interesting is that then they give you a way to try to define what PASC is in EHR data. Next slide please.

So both the study Suchitra spoke about and the work that I've been speaking about feed into the process of developing a set of rules or an algorithm also called a computable phenotype to identify which patients have PASC in EHR data. And there are two approaches that we've taken to this, both of which are ongoing. One is a rules-based phenotype. In other words, it's a set of rules based around specific lists of diagnosis codes and rules around when and under what circumstances those codes occur in order to say with some probability that a patient might have PASC. And additionally, we've employed a machine learning based approach which involves training a model to classify EHR data, which patients resemble those who are diagnosed with PASC based on input data that's selected from these tree based scan statistics selected features that we presented in this study. Next slide please.

So just briefly, our machine learning phenotype, the inputs it draws on are not just conditions but also procedures, laboratory testing and prescribed medications. And one advantage of it is that it outputs probability, so it gives you some level of confidence in its classification that a patient may have had PASC. Another advantage of machine learning methods is that they can pick up on very complex patterns that involve a lot of different specific inputs that are very hard to notice just by looking at the data. But the downside is it depends a lot on the data the model is trained on, and in particular the model we used was trained on both MISC and non-MISC PASC diagnosis codes. And if those codes are not representative of what PASC is, which there's good evidence to think they're not, this is a limitation that can be inherited by the model and we have to be aware of.

And for that reason, another approach we've taken on the next slide please, is developing a rules-based phenotype for PASC. So the rule set looks both at the specific diagnosis codes U09.9 and the MISC diagnosis code M35.81, which I spoke about initially. But the rule set also looks at patients who had COVID and then in the post-acute period, one to six months following acute infection, had diagnoses in one of several clusters of PASC associated diagnoses, which were put together by combing through the output of the two studies that you've heard about today.

And so, where we are with this phenotype is we're in the process of doing chart review validation, which means that chart review expert adjudicators are actually, for a sample of patients, looking at patient's charts and seeing whether when our algorithm says that a patient was probable to have PASC, whether the clinician who was seeing the patient agreed with that or whether they perhaps attributed the symptoms to something else that we didn't pick up in the EHR.

So we're in the process of summarizing the findings from chart review, and then next steps will include further refining this algorithm to better capture who has Long COVID in EHR data.

And finally, next slide please, I thought I would conclude just by pointing to a few other studies that are in progress, either in the form of pre-prints that have been submitted, or manuscripts that are currently being written or studies that are just being planned that build on some of this work. We've been working on identifying what are the distinct subtypes of PASC. We've been looking at exacerbation of existing chronic

conditions, such as diabetes and asthma and others, and we've been looking at temporal trends trying to see how has Long COVID changed over the course of the pandemic.

Additionally, we've looked at effectiveness of certain treatments and other topics that all interact with one another. So thank you so much for listening. I will stop here and looking forward to the discussion.

Christine Bevc:

Great. Thank you Dr. Lorman, and thank you Dr. Rao for sharing these complex findings with everyone today. And, we're going to dive into these questions here. Special thanks to those of you who have submitted those questions and with the time remaining, which we're about halfway through, which is great, we will be able to get to a number of those. But first I'd like to bring in our discussant, Dr. Tantisira, to give us a brief synopsis to connect these two presentations. And, you get first dibs also on the questions to presenters as well. Dr. Tantisira?

Dr. Kelan Tantisira:

Well, that's exciting. These have been absolutely tremendous talks and I appreciate both Dr. Rao's and Dr. Lorman's presentation. I did want to give a brief synopsis from my take on these two talks and as a respiratory epidemiologist, I found it very fascinating the way that electronic health records data can actually be modeled multiple ways from an epidemiologic perspective.

So for instance, Dr. Rao actually presented her data more in a prospective, or actually retrospective but incident cohort type of approach where she gathered everybody up front and they followed them over the course of time to really look at incident symptoms and conditions. That's really more of a traditional epidemiological cohort, and it's great to see that these types of studies can actually incorporate that.

And Dr. Lorman actually identified more of a traditional cases and controls upfront to say, "Here these are the individuals that we can identify with PASC and compare it to a match group of COVID positive individuals and COVID negative individuals in a five-to-one matched case to control. He did not present one thing that I would be fascinated by, potential risk factors in doing a traditional case control type of analysis, but I'm sure that he may have some of that data there.

And to further summarize Dr. Rao's presentation, in addition to doing this cohort type approach, she actually looked at 660,000 patients, 59,000 with SARS CoV-2 positive status at that point in time, and she reported a variety of interesting hazard ratios with top symptoms including changes in smell and taste and others with marked increase hazard ratios of greater than 1.5, including hair loss, chest pain and elevated liver enzymes.

And then, she also looked at conditions that were associated with PASC, including myocarditis, ARDS, myositis and mental health treatment, all with significant elevated hazard ratios of 1.5 or above. And then, she very nicely went through calculations for how many of these do we believe are truly incidents for PASC themselves as compared to other conditions by comparing incident proportion differences and detailing specific risk factors for the reported PASC, as well, including ICU admissions, age less than five and chronic complex conditions.

Dr. Lorman, as we mentioned, took more of a let's cluster these together, do a five-to-one matched cases to controls based on diagnostic codes and gave a really nice presentation on how to enhance the potential for those diagnostic codes into relevant clusters that may represent PASC longterm outcomes. He evaluated 1300 patients with PASC and did a five-to-one case control matched cohort design, and then he carefully reviewed the major systemic and syndromic findings providing diagnostic codes associated with PASC to try and highlight enriched populations.

And obviously as a pulmonologist and respiratory epidemiologist, I was fascinated by the fact that many of his findings actually did point to respiratory conditions, including one that I didn't know was enriched, which is obstructive sleep apnea. So, thank you for pointing that out to me as well. And then he further went on to devise how do we take this forward in terms of using machine learning and rules-based algorithms and concluded with other fascinating work. I love the idea of looking at disease subtypes for instance, and the effective chronic conditions on PASC.

So I do have a few questions for you guys upfront, and obviously we'll leave plenty of time for the audience to ask things. I'll start with Dr. Rao. You presented two slides looking at hazard ratios, one which was for symptoms and one which was for conditions, and there was some conditions and symptoms that seemed to overlap to me on just initial look at them. For instance, symptoms can include things like chest pain and

cardiorespiratory symptoms and respiratory failure, and those tended to be much more modest in terms of those hazard ratios, all 1.5 or less, as compared to conditions.

For instance, when I think as a respirologist, respiratory failure 1.16 and then you have something like acute respiratory distress syndrome and those hazard ratios are much larger, close to three, there seem to be some differences there. So, could you explain perhaps what elements of your study contributed to those reported differences between symptoms and condition?

Dr. Suchitra Rao:

Yeah, thanks for that question and just your astute reviewing of those different hazard ratios in those two groups. We were struck by that too, and I think some of it has to do with the way that some of these data are coded, or some of these diagnoses are coded and may just have to do with some of our analytic approach. So for example, when we looked at the conditions and we were calculating those adjusted hazard ratios, we did exclude those who had preexisting conditions in that seven day prior to the test and going back 12 months. That way we would just look at things that were potentially associated with having acute COVID versus an underlying comorbidity that you had.

And so that could be one reason to explain why we were seeing a little bit more of striking hazard ratios in the conditions versus the symptoms. There could also be something around the coding of some of these. So we know that coding is probably going to be a little bit more specific for conditions versus symptoms. Not all symptoms may be coded for different types of visits, so that those could be a couple of explanations.

Dr. Kelan Tantisira:

Great. And Dr. Lorman, I think this is actually more of something that you alluded to but I think is a key portion of the clarification of your approach. So, you mentioned that your models may be limited if the diagnostic codes are not representative of PASC itself, and we get to this idea of preexisting disease ... for instance, I see a lot of asthmatic patients in my clinic, and so if one of my patients would come in with worsening shortness of breath in the post-COVID era ... Have you assessed whether I as a pulmonologist in my EHR might be more likely or less likely potentially to call this PASC instead of relating it specifically to asthma?

Dr. Vitaly Lorman:

Yeah, that's a great question. I think there are multiple answers here that point in different directions. So on the one hand we know that patients with existing chronic conditions are more likely to have a PASC diagnosis in the EHR. So, that's one part of it. On the other hand, we also know there are a lot of patients both with existing chronic conditions and without existing chronic conditions who have PASC symptoms and conditions that are associated with PASC in the post-acute period who don't receive a PASC diagnosis. And I think in our work on phenotyping, on developing a PASC phenotype, I think one of the hardest things has been figuring out how to detect attribution, so how to figure out whether a diagnosis that occurs in the post-acute period, whether it's actually due to something that the patient had prior to having COVID, and even if it was something they had prior to having COVID, whether it was their underlying condition was exacerbated by COVID, which we would still consider part of PASC, or whether it was just their normal trajectory irrespective of covid.

Yeah, so attribution's really difficult. We're hoping that the chart reviews that we're conducting can shed more light on this. We're also using methods from natural language processing to look at note data to see cases where in the chart notes symptoms have been attributed or not attributed. And then I should also finally say in a pre-print, we did take a look broadly among patients with existing chronic conditions, their trajectories following COVID. So we compared a cohort of patients with chronic conditions who had a positive COVID test to those patients with chronic conditions who had a negative SARS CoV-2 test.

And we did find increased utilization for many different kinds of chronic conditions, but we think that there's a lot more to investigate there that depends on the specific conditions. So for example, we've looked at trajectories of patients with diabetes over time before and after COVID, and we've done a little bit of preliminary work with asthma, too.

So yeah, hopefully there'll be more in the future. But I think the short answer is that exacerbation of existing conditions and what's attributable to COVID and what isn't are some of the hardest things to disentangle in EHR based studies.

Dr. Kelan Tantisira:

So I have two additional questions before we open it up to a broad chat, and I'd like both of your

perspectives on these following questions. We just heard about trajectories and we looked at PASC as outcomes. Have you looked at more longitudinal assessments following a diagnosis of PASC to evaluate what the natural history of PASC might be?

Dr. Suchitra Rao:

Yeah, that's a really great question. I can start and then Vitaly feel free to add to that. I think it's a really important question. We're all really keen to know if someone is going to get PASC, how long do they have it for? Are we noticing any differences in kids compared with what we're seeing reported for adults? Can we glean anything from EHR data while we're enrolling our prospective cohorts into the other arm of RECOVER? So we did try to look at some of that in the work that I just presented to you where we were just looking at patterns or events over time and who still had an event after a certain period of time. It can be very challenging to do this type of work with EHR data. We've talked about some of the limitations.

I think that we are at the mercy of when something is present, when something is coded, then we can say, "Okay, we know this is present," but when something is no longer present, is that just because people are dealing with it and not seeking healthcare so we can't capture it or is it truly recovered? And so that's why we might have some limitations looking at that in EHR data.

So with that caveat, we did look at some of the time to recovery with our data set, and I recall that around that six week mark ... I'm sorry, actually the six-month mark ... after having a positive test, about 40% of patients still ended up having some sort of PASC feature. So again, I think there's more work that we need to do in this space, taking into account some of those limitations. Other reports in the literature talk about recovery for kids in general being a little bit shorter than in adults, but of course we all know and we've taken care of kids and have had experiences of children who have symptoms much longer than that.

Dr. Vitaly Lorman:

Yeah, I think just to add a little bit. So I think this is definitely an area where there are a lot of interesting questions, and I think there's a lot more we could do. In our work on PASC subtypes, we've been looking at in a patient utilization in each subtype before and after having COVID to see whether there are groups where there's a return to normal versus even after six months, whether patients still have increased utilization by different measures. So that's one thing we've looked at.

In patients who had diabetes, in kids with Type One diabetes, we've looked at hemoglobin A1C trajectories before and after COVID, so that's a pre-print, but roughly there the findings were inconclusive but if you look at just the plots of the trajectories, we see hemoglobin A1C values among patients who had COVID go up around the three-month period and then come back down around six months.

So, we have definitely been looking at trajectories. We've also looked at the history of PASC itself as an entity to see how manifestations of PASC have changed over the course of the pandemic. That's a study that's in progress, looking for example at the hypothesis, whether has PASC since Omicron become more ... has Long COVID, have its manifestations become more respiratory in nature.

Yeah, those are some of the things we're currently looking into but I think there's definitely a lot more to be done.

Dr. Kelan Tantisira:

Great, thank you. My final question for both of you is I think this is something many people are interested in, but in addition to who I am in terms of my position, I'm also one of the RECOVER main site PIs, and we're actively recruiting patients into the main cohort, and I think a lot of folks are really trying to ascertain what are we going to do with the main cohort data and how can we optimally use it as we start to look at it.

What are your thoughts about as we begin to analyze the data from the main cohort, how can we best use your EHR data to support and validate those main cohort findings or vice versa? How can the main cohort findings better inform the work that both of you are doing going forward?

Dr. Suchitra Rao:

Yeah, that's really a great question. Thanks for posing it. And, I do agree, I think that there's a lot of opportunity for us to be cross-collaborative and that the work that we do can inform the prospective work and vice versa. And so, I think there's a number of ways. I think one of the objectives that we're really going to be focusing on in some of our future work is this concept of help with trial recruitment. There is overlap with the EHR sites and the prospective trial sites, and so there's an opportunity for us to develop, validate a computable

phenotype where we can identify patients that have PASC based on what we have seen and gleaned from the EHR that we can use then as a population that can be screened to see if they do indeed have symptoms of PASC, ongoing symptoms of PASC, and then could potentially be recruited into some of these studies.

So I think that's something that people are really wanting to pool resources and look into for the upcoming year. I think that something else that can really be helpful is I think of EHR data as a great way to do hypothesis generation and signal identification. So if doing some of these machine learning, unsupervised analysis, if we do find signals in the data, then that is something that can then be validated or measured or explored further with the prospective cohort. So it might inform study design, it might inform what type of patients should be enrolled. So, I really think that that has been super helpful. I'm looking at some of this in my work where we might find a signal and then I now want to explore this in my other work looking at the mechanistic aspects of Long COVID in kids.

I think another thing that we could potentially do is we have such a depth, there may be other conditions that we haven't talked about today that are presenting in children. Someone had mentioned neuropsychiatric findings or tinnitus, or some things that are reported that have been seen, but if we want to explore that further, EHR data sources are a great way to try and find those patients. And then, we can again help to figure out is there an opportunity for them to be enrolled in trials or are there specific questions that we can answer now that we've pulled and found a cohort of patients to be able to research further.

I think validation is something that you talked about. So again, we are doing a lot of work around generating computable phenotypes, looking at different sub-phenotypes, so the hypothesis generating and some of that work that we're doing can be validated with some of this prospective data. That's just a few. Vitaly, I'm sure you can think of others as well.

Dr. Vitaly Lorman:

Yeah, thanks Suchitra. I completely agree with your excellent answer. Yeah, certainly the HER cohorts give us a great platform for emulation for testing hypotheses before they turn into trials for the prospective cohorts. One thing also worth mentioning is that we have large sample sizes, so we might be powered to find statistically significant effects that might not be detectable in the prospective cohorts, and then vice versa, as Suchitra alluded to in our computable phenotyping work, these questions of attribution of how to define exacerbation, the survey data from the prospective cohorts linked to EHR data could be a great place to look into some of those questions that are hard to answer from the discreet EHR data alone.

Dr. Kelan Tantisira:

Great. I will hold the rest of my questions for now, although I have some ready to go because there's so many great things to ask, but I do want to give an opportunity for the questions in the chat to be answered.

Christine Bevc:

All right, well thank you Dr.Tantisira. If we get a few minutes at the end there to circle back, we'll see if we can squeeze those in. We do have a lot of questions that have been coming in. Some of them have already been answered by our panelists. Dr. Rao dropped one into the chat there. That's moved over under the answered questions, so you can check and see if your question has already been answered, but please continue to bring those in there.

So, the first question I have continues what we were just touching on about the value of the HER cohorts and testing hypothesis. This is a question to both our panelists about whether you've been able to look into the effects of possible interventions, such as patients that may have been prescribed pharmaceutical interventions. We could name them, but those that are out there as possible interventions for Long COVID. Is that something that has been talked about? That you're aware of that it's on the list of future things to do? I wonder if you guys could speak to interventions and cohort data.

Dr. Vitaly Lorman:

So we're in the process of working on a study looking at specifically Remdesivir and Dexamethasone, their effect as COVID treatments on risk of developing Long COVID. We've looked at Paxlovid a little bit in a query in the past, and I think there have been other treatments also discussed. So there's definitely work in progress on a few different treatments. I think also in the study that Suchitra represented on, we did also take a look ... this is in the supplementary appendix ... at broadly which drugs tend to be prescribed to COVID

positive patients as compared to patients who don't have COVID. So not effectiveness so much, but just looking at in practice which drugs tend to be prescribed.

Christine Bevc:

Dr. Rao, anything to add to that?

Dr. Suchitra Rao :

Yeah, I was just going to say, so then based on that work that we did in that was an adjunct to the work that I just presented, some of those medications were really used to try and treat some of those PASC symptoms. And so it was more just those supportive care measures, so if folks had ongoing cough, it was looking at cough suppressant medications or non-steroidals and things that might be helping, again, more of those acute symptoms.

So again, we just saw that there was some association but there wasn't enough power to really be able to show that that then led to improved recovery or faster recovery so that those weren't really designed to look at true comparative effectiveness.

So the only other thing I was going to mention is really the work that we've done in the first year of this RECOVER initiative is to look at detecting PASC and defining PASC, and then the next steps for us are really going to be looking at treatments of PASC. So speaking of the work that Vitaly is talking about, that's definitely going to be a very heavy focus for us in the next year, as well as looking at preventative aspects as well. So are there other things that we can do to try and prevent those symptoms from occurring in the first place, or at least reducing those symptoms duration?

Christine Bevc:

Great. So the next question we have actually is coming ... It's been asked a couple times by our audience members, and this gets into some of the rules set and the inclusion and the identification of patients within the dataset. How is all of this impacted by those that may not have had access to testing early on? So in spring of 2020, their being able to get to tests was difficult or later on where they may not have tested because symptoms might have been more mild and then complications developing further? So I wonder if you guys could speak to the absence of testing and how that plays in?

Dr. Suchitra Rao:

Yeah, I mean the folks that posed these questions absolutely hit the nail on the head that we're really at the mercy of the data that we have that come in to EHR, and we acknowledge that that is going to be one of the limitations of looking at things like the incidence of PASC at the beginning phases of the pandemic. So we know that testing at that time was very much restricted to sicker patients, those who might be more likely to be hospitalized, had access to those test centers that were built around those larger facilities. We definitely acknowledge that, and have seen a little bit of a signal for things that may not really be the case and so that's what we have to take into account. And then, testing became more widespread and more reliable.

And then we went into the issue of at-home testing. So now people aren't going into those centers to get the tests, they're doing all of their testing at home. So the period of time that I looked at with our study was March of 2020 through to October of 2021. So we're less at the mercy of the at-home testing phenomenon, but we certainly are at risk of bias when we are looking at that early period. And so when we design studies, we have to think about these things. It's still better to try and explore these questions using what we have, but there may be some caveats. We might not be able to give those precise incidence estimates at the start of the pandemic for example, for those very reasons.

And then we have to think about a combination of approaches. So it made sense to really look at testing during that first period of time when it was more reliable. But then when you get into latter phases of it ... Oh, and we couldn't use the U09.9 code because it didn't exist until October of 2021. So then you have to evolve the way you design your studies. And so maybe we use more of reliance on other codes, or you have to acknowledge that there may be limitations with looking at SARS CoV-2 PCR testing alone. So that's where over time we've had to take into account these factors, look at a combination of different approaches and this is why it's good to do work in different ways because we can then look at similarities, look at differences, look at if something works, then pursue that technique further. So, that's a really important and great question.

Dr. Vitaly Lorman:

Yeah, totally agree with everything that Suchitra said. One specific issue that we've discussed a

lot has been that patients who are diagnosed with Long COVID with that U09.9 code, more than half of them don't have a prior viral positive test result recorded in the EHR. So, it may be one that was conducted at home or outside of the network which feeds into the EHR data. And so that makes it challenging, for example, to figure out the time at which patients had COVID, which is important for some studies.

There are some things in the case control comparison studies that we've done to try to mitigate this somewhat. For example, oftentimes our control groups have been patients with a negative SARS CoV-2 test and using that to determine the index date. So at least there's some more confidence that the patients didn't have COVID at the time of the index date, even if we didn't capture the positive COVID test results at other times.

There's also been some studies where we've looked at non-contemporaneous control cohorts such as patients with respiratory illness in 2019, so that the presence of patients with COVID in our control cohort where that's not a concern, although that comes with its own difficulties concerning differences in utilization prior to the pandemic as opposed to during the pandemic.

Christine Bevc:

A related question to this one is with the ICD codes that are used, thinking more broadly, what about international data? Are there other countries that have Long COVID codes or is this an approach that could also be replicated in other EHR datasets across the world?

Dr. Suchitra Rao:

Yeah, that's a great question. So we do know that there are some countries that do have diagnostic codes for Long COVID and Long COVID conditions. I don't have a list of all of the different countries at easy access, but that does afford some opportunity for us to take a similar approach and explore that in another setting.

But it's very across the board, they're instituted at different times, and then it's just a good way to highlight some of the challenges of trying to do cross-country collaborations because again, there might be different utility of when those codes are being used in those settings compared with our settings, et cetera.

Dr. Vitaly Lorman:

Yeah, unfortunately I don't know too much about coding practices in other countries. I think that's a great question, but I think outside of my expertise.

Christine Bevc:

All right. Well, thank you. So our next question, we'll start with Dr. Rao and then come back over to you, Vitaly. So in the analysis that the team's looking at, how do individual symptoms like tinnitus or sleep disruptions, how do they factor into the study? Recognizing that there is a wide range of symptoms, what about those symptoms that might be less common or those that we didn't see on the slides listed today?

Dr. Suchitra Rao:

Yeah, thanks for that question. So, absolutely. So there is such a wide spectrum of manifestations of PASC, and you saw in those earlier slides that based on even just review of the literature, clinical adjudication, sharing our experience in terms of taking care of these patients, we had 441 clusters of potential diagnoses of symptoms or conditions associated with PASC. So, we didn't report out all of them. We really just focused on the ones with the strongest association. But yeah, absolutely there's going to be other conditions and other findings that have been reported and new ones that might be continuing to be reported.

So those do factor in when we're trying to look into incidents rates of PASC. Those become really important when we're trying to characterize the full spectrum of PASC, and those may then be relevant too when we're doing prospective studies. We make sure that we're asking questions about things like tinnitus, about things like different memory problems or skin rashes. So, I think these then get factored into other studies down the road.

The study that we were doing was really trying to look at those stronger associations, but we acknowledge that there's that whole spectrum of things that may not necessarily be flagging in the highest risk category.

Dr. Vitaly Lorman:

Yeah, just to add to that, I think this is one place where the tree based scan statistic, that approach might provide some additional signals. So for example, I did look to see whether there were ear infection related codes that occurred disproportionately among the cases in our comparison, and when we compared

patients with COVID to patients without COVID, so that sensitivity analysis I described, there we did see some signal for some ear infection related codes. I think that's the value of doing multiple comparisons and trying to synthesize them.

And yeah, I think in our manuscript there's a table that lists clusters of specific codes that all seem to be associated with PASC diagnoses or with COVID positivity in the post-acute period. And so going through these in more detail, having studies specific to those conditions down the road we think would be really important.

Christine Bevc:

Great, thank you. And for those of you who are on the line, for all the mentions about the publications, we have a publications page on RECOVERCOVID.org where these prints will be available and linked to so you can check those out. And then also, in the survey questions, if there's any questions or ideas or topics you'd like to see in the future, you can jot those down there as well.

We've got time for one or two more questions. This is going to be an easier one here. Or, at least I think so. How do all of the different variants of COVID-19 play into this analysis? We've heard Omicron mentioned, and then we also have early on with the severity of the different variants. How does that tie into some of the analyses that you guys are looking at?

Dr. Suchitra Rao:

Yeah, another really great question. So when we think about the whole course of the COVID-19 pandemic, we do need to be thinking about these in terms of different waves, but we have to be careful not to assume that the patterns that we're seeing are related to a variant and they're inherent virological properties. It could be other factors too. So it could be that over time we end up with a population that gets immunity through exposure from infections and from vaccination. With EHR data, there's also changes in people's practices in terms of how they code data, in terms of new codes that develop. We talked about U09.9. And so, when we conduct these studies and do these, we have to think about not only the inherent properties of the virus, the changes of immunity over time, the ways that the healthcare system has been managing these with new therapeutics, et cetera, but also how patterns of healthcare seeking and healthcare reporting change over time, too.

And so, with the study that we looked at, again, we were looking at the early phase before the Omicron period hit, and then Vitaly's study looked at periods of time beyond that. And so we do a lot of work where we try to break things down into phases too and see, okay, during different time periods are we seeing differences? But we have to acknowledge that it may not just be because of that variant that we're seeing these signals.

Some other work that you've heard Vitaly talk about is temporal trends, so looking at different presentations of PASC and how they change over time. We've been doing that over a variant period. We did something similar with MISC. So we wanted to see what was happening with the severity of MISC over different time periods. We found that the severity was actually going down, but that's not to say that Omicron is less virulent necessarily, it's just there may be more population immunity.

So, just to summarize my response, it is something that we are looking at with a lot of the research that we're doing here, but we're trying to be mindful of what some of those changes and variations across the time periods truly represent.

Dr. Vitaly Lorman:

Yeah, I second everything Suchitra said. Just to give one more example, in our study, in our pre prints on kids with diabetes and their trajectories over time, we did look a little bit at pre-Omicron and Omicron effects and we saw potentially some differences there. But as Suchitra said, it's not just the variant, there's potentially a lot of confounding from just utilization patterns. Like in summer 2021 when patients started coming in for other things more, that can throw a wrench into a lot of analyses and require more careful planning for sure. As well as other respiratory illnesses, flu, RSV becoming diagnosed more frequently, potential for confusing symptoms of that with COVID, all of those things become more difficult later in the pandemic.

Christine Bevc:

All right. And we're going to close with one last question, and really this is circling back to one of

the earlier questions about the comorbidities and risk to children and adolescents with preexisting conditions. And I wonder if you both could speak to what you're seeing within these populations and whether there's disease types or other confounding factors that may be increasing the risk of seeing PASC in these children?

Dr. Vitaly Lorman:

Yeah, so maybe I'll start just with we did a study that just took a broad look ... The one I mentioned before, comparing patients with existing chronic conditions who had COVID to those with chronic conditions who didn't. And what we saw there is I think we saw ... I don't remember off the top of my head, but something like patients with chronic conditions were somewhere around twice as likely to be hospitalized following COVID infection compared to patients who had chronic conditions who did not have COVID. However, when we looked into specific kinds of chronic conditions, we saw effects across a lot of different clusters and it wasn't so clear to us and it's hard to disentangle how often there's a causal effect, there's some biological mechanism between COVID infection and something that exacerbates the chronic condition versus just increased surveillance, like patients who have a chronic condition and then have COVID being more likely to see the doctor, be admitted to the hospital just by virtue of having the chronic condition and COVID at the same time, but not necessarily there being a causal relationship.

So those things are very hard to disentangle. And in our work on Long COVID subtypes, we're definitely seeing some differences between COVID subtypes based on patients who have presence of existing chronic conditions. There seem to be subtypes involving patients who have high volume of utilization, have chronic conditions affecting multiple body systems, have a lot of laboratory testing. And to the extent to which it's the existence of the chronic condition that's driving this versus something about how PASC interacts with this is definitely a challenging question that we're trying to look into to differentiate these further.

Christine Bevc:

All right. Thank you, Dr. Lorman. And with that, please join me in thanking our panel of investigators and our discussant for taking the time to share these updates with us, and our audience for joining us today. The FAQ for this webinar is going to be posted along with the recording on RECOVERCOVID.org. The FAQ is going to include the answers to the questions that are relevant to today's webinar, as well as those that were submitted in advance or during the session.

Questions about other scientific topics will be addressed in future webinars, and answers to broader questions about RECOVER will be available in the general FAQ section on RECOVERCOVID.org.

As we close, we invite you to come back and attend future R3 webinars as we dive deeper into some of these broad topics that were discussed today. If you haven't already, remember to sign up on the website to receive future announcements and details about our upcoming webinars.

And lastly, if there's topics that you'd like to learn more about, please be sure to enter those ideas into the survey that's going to be popping up onto the screen and write those in there so that we can make sure that we're sharing the results back that are of relevance to you.

Thank you again to our presenters and our audience for joining us, and this concludes today's R3 webinar. Thank you.