Transcript

Quinn Barnette:

Welcome everyone to the Recover Research Review or R3 Seminar. My name is Quinn Barnette. I'm an epidemiologist with the Recover Administrative Coordinating Center and I'll be your moderator for today's session. The goal of this seminar series is to catalyze a shared understanding of the research within the Recover consortium. I want to start by thanking everyone who submitted questions in advance. I remind everyone that you can submit any questions during today's presentation using the Q&A feature in your Zoom menu. After today's [inaudible 00:00:32] Q&A document will be posted with the recording of the seminar on recovercovid.org. The document will include the answers for submitted questions relevant to today's presentations. Questions about other scientific topics will be addressed in future seminars and answers to broader questions about Recover will be available in the FAQs found at recoverCovid.org. As a reminder, we cannot answer questions about individual clinical care.

I'm pleased to share that our presenters today are Dr. Richard Moffitt, Ms. Emily Hadley, and Ms. Hannah Mandel. And our discussant will be Dr. Ravi Jhaveri.

Dr. Moffitt is an associate professor of hematology and medical oncology with a secondary appointment in biomedical informatics at Emory University. He trained as a biomedical engineer at Georgia Tech where his early work led to the discovery of subtypes of pancreatic cancer from gene expression data. As part of the leadership team for the National COVID Cohort Collaborative or N3C, he's helped develop a community of over 4,000 registered users working with health record data of over 20 million US patients aggregated from over 200 health care sites.

Hannah Mandel is a senior research scientist with a clinical science core at NYU Langone. Her work focuses on leveraging electronic health record data to support population health and clinical quality improvement. She received her master's of science in Biomedical and Health Informatics at the University of Washington and has spent the past decade working as an informatician in New York City joining NYU in 2022. She's particularly interested in data quality initiatives to facilitate secondary use of healthcare data.

Emily Hadley is a research data scientist with more than six years of experience of RTI International. In her work, Emily collaborates with subject matter experts to solve complex problems in health, education and criminal justice using AI data science statistical approaches. Emily has led the Recover RTI data science team for more than two years and has contributed to research related to long COVID and pregnancy as well as the past computable phenotype. Emily is passionate about the responsible use of AI and leads RTI's contributions to the US AI Safety Institute Consortium.

Dr. Jhaveri is a division head of pediatric infectious diseases at Lurie Children's Hospital and professor of pediatrics at Northwestern University. His research focuses on hepatitis C, particularly on infants, children and pregnant women. He serves in the American Association for the Study of Liver Diseases, HCV Guidelines Panel and Viral Hepatitis Elimination Task Force. Dr. Jhaveri is a fellow of the Infectious Diseases and Pediatric Infectious Diseases Societies and Chair of the Infectious Diseases Society of America's Standards and Practice Guidelines Committee. The topic of today's seminar is

Patterns of Past with Initial COVID-19 Infections and Reinfections EHR Insights. Today's speakers will present research findings about the instance of long COVID in adults and children between 2020 and 2023, focusing on changes over time. They'll also discuss the risk of developing long COVID following first COVID-19 infection and reinfection. Please welcome all of our speakers, and with that, I will turn it over to our first speaker, Dr. Jhaveri.

Dr. Ravi Jhaveri:

Thanks, Quinn, and good afternoon everybody. Good morning to some of you, depending on where you are in the country, but I'm very excited about the session. I think that one of the really important questions that we are still learning the answers to are what are the risks of PASC in the current environment. So much of the data that we've generated thus far really was derived from the early years of the COVID pandemic, the original strains and the Alpha through Delta strains. And I think so many of my patients ask me in our clinic, well, what is the risk of long COVID currently with the current variants and the like? And so I think our group of speakers today have been really trying to answer some of these key questions. And so as I said, I'm very excited about the discussion today. And so I'm going to turn it over to our first speaker.

Dr. Richard Moffitt:

Thanks, Ravi. Emily, are you bringing up the slides? All right. So, Emily and I are going to start by presenting work on a paper that was published earlier this year, insights from an N3C Recover EHR-based cohort study characterizing SARS-CoV-2 reinfections and long COVID. And while Emily and I are here today, we really represent a very, very large team of researchers from N3C and Recover. Next slide, please. So as Ravi alluded to, most of our initial work was based on primary infections, but we know that as more and more people are exposed to COVID, more and more folks are getting reinfected with COVID. And so what we wanted to do was see what insight we could learn from electronic health records in patients that had had reinfections, either looking at their first infection, their reinfection or subsequent reinfections.

Next slide, please. So as I mentioned, this is an electronic health record based study and the good news is we have over 20 million patients worth of EHR data coming from more than 60 healthcare sites across the US. It's important to note that this cohort of folks is not the Recover observational cohort, which is being followed particularly over time. And so as a result, we only have information that is collected as part of medical care. So this includes all the diagnoses these patients would have, including relevance to this work, COVID-19 infection itself, as well as PASC or long COVID diagnoses that they may have. We have access to the lab test results and vitals for these patients, including of course COVID-19 tests as long as those tests are included in the medical record in some way. We also have indications for medication including Paxlovid so we can infer COVID incidents from Paxlovid.

We also know the visit type, so we can tell if it's an outpatient visit or an inpatient visit or an ICU visit, which is what we use to define the severity of disease in this setting. What it does not include is undiagnosed conditions, especially those that are coming from home testing. So if your doctor doesn't know about it, it's not going to trickle back to us, which means that most of the word here needs to be viewed through the lens that we are missing a lot of asymptomatic or mild infections that never rise to the level of importance for that person to seek care. Next slide. So in this study, we ended up with a final

cohort of patients, 212,000 of which were reinfected from a larger group that included 2.8 million patients with solid evidence of primary COVID infection, but no evidence of reinfection at least as far as the dates of this study.

And so we can start by looking at the difference between the patients who have reinfections and those who we have no evidence of reinfection and we see that folks with reinfections are younger, they're more proportionately Hispanic or Black, they have higher Charlson Comorbidity Index scores, which suggests they have more pre-existing conditions and they're also less likely to have been vaccinated prior to that first COVID infection. So at this point, I'm going to hand it over completely to Emily to take us through the rest of the results.

Emily Hadley:

Awesome, thanks so much, Richard. Hi folks, I'm Emily Hadley. I am a research data scientist as was mentioned and I'm excited to share the results from this work. As Richard mentioned, this was a huge team effort and so many of these results were from a variety of different contributors to this work. We're going to start with the incidence results and summarizing how frequently do we see reinfections in this data and when are they occurring? This figure here is showing the percentage of persons with reinfections at various points in time. We have divided the figure itself into when the reinfection occurred and then colored these particular plotted lines by the initial infection epoch. And so you can see that we are tracing from the ancestral time period through later Omicron periods in 2023, which was when we kind of wrapped up this analysis and put together the initial draft of our publication.

We observed the largest increase in reinfections in the Omicron BA.1 and BA.2 epoch with an overall incidence of 6.9%. And that 6.9% is pretty similar to what we've seen for incidence in other large EHR cohort studies in the literature. Our second finding related to incidence comes from this comparison heat map where again, we're looking at when the date of the initial infection epoch was with the first reinfection epoch for an individual. We can see here that the reinfections in the Omicron BA.1 and 2 time periods were particularly common among individuals infected in the ancestral COVID-19 and alpha, beta, and gamma time periods. This aligns with other work that we've seen in the literature which suggested a pretty significant rise in reinfections as well as initial infections in the Omicron period as a result of that particular variant.

And then the third piece that we looked at when we were looking at incidence was the time to infection. And here, we can see a plot of the time between the initial infection and the reinfection based off of the initial infection epoch. Our reinfections needed to incur at least 60 days after the initial infection, which is why you can see that marker on the plot. And our key takeaway here was that in many cases, reinfections were occurring more than 100 days after initial infection and sometimes quite far out. We documented 156 persons in our data that had a reinfection more than 1,000 days after their first infection. This was notable at the time we wrote it up because there was not really a strong quantitative understanding around reinfections and their timing. I think this past summer in particular, there was a lot more dialogue around reinfections in the news media and awareness that you could get a reinfection and perhaps not just once but multiple times. And that's what our data suggested as well.

The next piece that we looked at in our study was biomarkers. So we wanted to understand what the relationship was between various biomarkers and reinfection. And our key takeaway here was

specifically related to the biomarker of albumin. Albumin levels were persistently lower after the initial COVID-19 infection and remained low leading up to a reinfection. We didn't necessarily see this trend in some of the other biomarkers that we looked at. Some of our key findings overall were related to severity. This is a table that compares the severity of initial infection on that Y-axis with the severity of the first reinfection along the top or on the X-axis. And we're looking at each box represents the count of persons who had a reinfection. And then as Richard mentioned, because we have information around visit type, we can classify the severity of that infection. So our categories here are mild, which means that there was no documented emergency department visit or hospitalization.

We had mild with an ED visit or mild with an emergency department visit. Moderate was some degree of inpatient hospitalization and then severe was hospitalization that required the use of ECMO, IMV or vasopressors. And then we also had patients that passed away as well. Our main takeaway here is that the large majority of persons with a reinfections had both mild initial cases and mild first reinfections as well. But for patients who did have a severe first infection, more than half of them experienced a severe reinfection that required either a visit to the emergency department or inpatient hospitalization. And so we would like that takeaway to be that if you had a severe infection the first time, there may be a possibility of a severe infection the second time as well and you should pay attention to that possibility.

And finally, we started investigating the relationship of reinfections and long COVID. This was really meant as an observational understanding to see what can we learn from our rather large data set about the relationship of reinfections and long COVID diagnoses. And what we found was that the incidence of new long COVID diagnoses after reinfection appears lower than after initial COVID-19 infection. And you can see that in the figure because all of our dotted lines which indicate reinfection are above the solid lines which indicate diagnoses after first infection. This is where we are particularly interested in doing some follow-up to understand at a deeper and causal level what might be related to this finding.

We have some limitations of our work. The first is electronic health record data has a variety of limitations for research purposes. One of them is a selection bias towards more severe events where people are seeking out healthcare. Another is related to limitations for access to care. Richard mentioned to this earlier, but we don't include home tests as part of this data because we don't capture them, which might contribute to an underestimate of our reinfections. Most of the analyses we just discussed here are limited to only the first reinfection, so the first infection you have after your initial infection. We do have persons in the data who have evidence of multiple reinfections, which might be intriguing as a future follow-up study.

Our inclusion criteria permits only a COVID-19 diagnosis rather than requiring a PCR or antigen test for the initial infection and only a PCR or antigen test can be used to document a reinfection. And I'll note that we do have some sensitivity analyses in the paper that explore those two criteria. Our next steps are that we are continuing to refine and better understand the patient cohort and we have continued to do ongoing analysis with reinfection, specifically focusing on understanding a causal relationship between reinfections and long COVID. We're preparing that work now and are hoping to get it out in the near future. That is it for me. I believe I am passing it to Hannah now.

Dr. Richard Moffitt:

Yeah, and I think we're going to hold to answer your questions until we're finished with both presentations. So I'm going to start off this next one as well on behalf of Hannah and myself and another very large group of researchers that have collaborated to study long COVID incidents in adults and children. Next slide, please. So whereas the previous talk was all data from N3C in one particular EHR cohort, this work represents a wide collaborative effort across three different EHR repositories, the N3C repository supported by NCATS as well as PCORnet adult and PEDSnet supported by PCORI all together, all supported by the Recover program. There are some subtle differences between the cohorts. N3C is large, 84 sites, 23 million patients. They have linkages to death records claims and geocoded environmental data. Similarly, the PCORnet adults has all those same linkages with 41 sites and 20 million patients, very comparable in size and scope.

And the PEDSnet group focuses on specific children's hospitals within those same networks. We go to the next slide. This will give you an idea of the geographic diversity. So we don't claim to be completely representative of the entire US, but we are geographically diverse collecting data from patients across the US in a variety of settings. While there is a bias towards the large academic medical centers, we do include data from more community organizations and have worked very hard to include data from patients in rural settings as well as urban settings. Next slide. So the overall goal of this work that we're going to talk about is to estimate the prevalence of long COVID and really to ask questions about how that incidence has varied over time and sort of have a unifying perspective on why this incidence estimate may change across various studies and change over time.

In this work, we looked at the incidence among both adult and pediatric populations in the three cohorts that I described earlier. And what Hannah is about to tell you is about all the different temporal patterns that we observed across these networks. We've used different definitions of long COVID, so we're really just trying to see all the different ways that we can estimate incidence over time. And finally, identifying key associated factors that may be driving these trends. So I'll hand it over to Hannah. Thanks.

Hannah Mandel:

Thanks, Richard. Yeah, first, I'm diving into some methods. So for our patient population, we started off by identifying patients who had indication of an acute COVID infection in their medical records. So whether this was a positive test, a diagnosis code, or a relevant medication, we selected the first COVID infection documented for each patient and restricted to those taking place before February 2023 to allow all patients minimum of six months to assess whether the patient developed long COVID. And we also restricted to patients who had at least one visit within the health system prior to their first infection and at least one follow-up visit at least 90 days after. And so that patients were more likely to be accessing care within the healthcare system and have clinical data available. Next slide.

So how did we define long COVID? Each of our networks has for the past several years, been working to develop their own long COVID definition for either adults or children. And they range from broader and more inclusive to more restrictive. But given the trade-offs between sensitivity and specificity with different definitions, we really felt it was appropriate to examine a range instead of arriving at one single estimate and saying that is long COVID incidence. So each network applied their

algorithm or we also call these computable phenotypes to see which COVID positive patients developed long COVID within 180 days. And we focused on the patterns that were evident across networks. There's a few details on these algorithms here, but just briefly, N3C's approaches machine learning based and it was actually a focus of the last R3 presentation. It's trained on a set of patients with a formal long COVID or U09.9 ICD-10 diagnosis code and predicts which other patients have on COVID based on their clinical and healthcare utilization data.

And then PCORnet and PEDSnet have rules based algorithms that mainly identify patients based on the presence of new onset diagnoses found to be more common in patients with COVID or long COVID or distinguished based on clinical input. So just a quick caveat, this outcome would be more accurately termed probable long COVID, but I'm just going to be calling it long COVID from here on out for simplicity. You can move to the next slide. So our analysis really focused on incidence proportion or the percentage of COVID positive patients we identified as having long COVID within 180 days of their acute infection. And we looked at other variables such as age, sex, patient race, ethnicity, the rurality or whether the patient's location for their clinical record was in a more urban or rural setting.

The variant era based on which variants were dominant across the US at certain time points, documented vaccination and burden of preexisting conditions, which I'm considering here to be the CCI or Charlson Comorbidity Index for adults and the patient medical complexity algorithm for children, sorry, and I skipped severity of acute infection. So we base that on, as Richard alluded to earlier, evidence of infection documented during a hospitalization with indication that it was for their COVID versus being incidental or diagnosed during another type of visit like an outpatient or emergency department visit. And we also present hazard ratios for risk of onset of long COVID.

We can move on to the results. So the majority of our patients were white and were female. We had the majority of patients with mild infection, meaning that their COVID was diagnosed during an outpatient visit and the percentage of COVID patients categorized as developing long COVID really ranged. So for adults, between nine and 26% of patients were classified as having long COVID. And for PEDSnet 7% of children were categorized as developing long COVID. Patients with long COVID were more likely to have been hospitalized for their acute infection and to have a higher burden of pre-existing conditions than those without. We can move on. So now, we're looking at monthly counts and proportions across networks. So on the left, we see on the top on COVID, which is labeled as past case counts and below that COVID case counts over the study period. And we see different variant errors identified by dotted vertical lines with Omicron divided further by subvariant.

So you can see that counts are lowest in PEDSnet, which is in green, and that's true for both COVID cases and long COVID cases. But overall across networks we see similar patterns. So counts are quite low around March of 2020 and then we see a spike in COVID and subsequently long COVID counts at the beginning of the Alpha period and another bigger spike at the beginning of the Omicron period. And just noting that the data at the beginning of 2023 shows a decline in long COVID cases, but that disappears once we add in more follow-up data. So it's likely an artifact of incomplete follow-up times for some of our population due to the varying reporting schedule of our sites. So this paper is in pre-print and we're revising for resubmission. That's one of the things we'll be adjusting. And I've indicated that kind of more uncertain period in blue.

On the right, we see the proportion of COVID positive patients developing long COVID within 180 days. And you can see that PCORnet, which is in yellow has the broader definition and we see a

higher proportion. But even with the different definitions, we see the same patterns around the variant periods and overall a fairly stable incidence over time. On the bottom right, you can see trends in percentage of COVID positive patients that were severe, treated with Paxlovid and had evidence of prior vaccination over time and just noting that the proportion of cases with prior vaccination and with prior treatment, I mean, treatment at the time of acute COVID are lower for the pediatric population as we'd anticipate, but there's a lot of concordance there across the adult population.

Moving on. So the first item we're looking at here at the top is a heat map displaying incidence proportion, but this is by COVID severity going along the Y-axis. And at the bottom is patients who were not hospitalized at the time of their COVID index infection to at the top hospitalized with intensive care, and across the X-axis for each network, we see increasing burden of pre-existing conditions. So that is from increasing from left to right for each cohort. And incidence proportion goes from 0%, which is blue to a 100%, which is red, with white representing the overall incidence for that network. And you can see generally as severity increases and as burden of pre-existing conditions increases, you can see a trend towards higher incidence proportion. And finally below that, we have monthly relative hazards, and that's for risk of long COVID. We're looking at that monthly and we can see unadjusted in black and adjusted in orange. Again, we see patterns similar to what we looked at on the prior slide with incidence proportion over time and a lot of concordance again across our networks.

Next slide. So finally, we performed a separate analysis to examine long COVID incidence among COVID positive patients as well as negative controls and pre-pandemic or historical controls. We anticipated because long COVID symptoms are so wide-ranging, and not necessarily specific to SARS-CoV-2, we would see a baseline or background level among both control populations. And we also included both since they both have different biases, which I'll explain once I get to our limitations. So we started off by identifying COVID positive patients and we looked for positive tests during the first five months of 2021. And our logic there was testing was available, it was before widespread home testing, and also those months were all part of the Alpha variant, when the Alpha variant was dominant.

We also looked at control groups that were negative, so having a negative test during the first five months of 2021 and patients that had visits over the same period back in 2019. And after doing that, we took things a step further, we estimated excess incidence. And to do that, we looked at long COVID among our COVID positive patients and subtracted out that background level. We tried to get as close as we could to subtracting out what background level of long COVID symptoms we would see among a healthy population. For us, the closest we could get was our outpatient negative controls. And yeah, this is a very conservative estimate of excess incidence given that outpatient negative controls are still getting tested for a reason and may be sick or having false negative tests, but we considered this a lower bound estimation.

And what we found was that we did see some level of long COVID-like sequelae in both of our control populations, but at lower levels than among our COVID-positive patients. And that our excess incidence ranged from 4 to 7%, so 4% in children and 4 and 7% among our adult populations. We can move on to the next slide. So just to reiterate some of our key findings here. The long COVID incidence proportions for adults and children ranged, we saw a lower bound around 4 to 7%. The majority of patients did not have a severe acute illness. However, the severity of the acute infection was the strongest predictor we found for subsequent development of long COVID. We note that patients hospitalized around the time of acute COVID may have more clinical data available and be followed

more closely, so they would have more of an opportunity to be flagged as potentially having long COVID either by a clinician or one of our identification algorithms.

And finally, we found that patterns in incidence over time and risk factors were largely concordant across our cohorts. We didn't see sizable decreases over time, although recent publications, some of them suggest risk is lower with more recent variants, some haven't seen that. And although changes in variants and introduction of treatments, actions and vaccination could be lowering on COVID risk over time, there are also a lot of other temporal factors that make peak coming into play. So there's increasing recognition and documentation among clinicians rebounding healthcare utilization since the beginning of the pandemic, potentially cumulative impact of having multiple infections, which we did not examine here. So even with our conservative estimate, it's clear that ongoing COVID infection still presents a large burden. Next slide.

So some of our limitations overlap with the ones Emily presented. EHR data represents patients who sought healthcare, so patients who are too sick to come in or did not have any follow-up visits or were less able or willing to access healthcare would be underrepresented as well as patients in rural settings or who received care at non-academic health centers or impacted by other structural or social determinants of health. We also had incomplete vaccination data, and this is partially because many vaccinations were distributed in non-clinical settings, so it's hard to draw conclusions about the role of vaccination played in the results. Finally, diagnoses of COVID, acute COVID infection and long COVID may also be missing in electronic health record data for a variety of reasons. So this includes early unavailability of testing, the introduction of home testing, delay in having standardized codes to capture long COVID and some of its manifestations like POTS, and then complexities and diagnosing long COVID at the point of care.

And as mentioned earlier, there are some limitations around our control groups. So negative control group patients may have had false negative tests and could actually be a less healthy population depending on why they're actually getting tested. And then while we know patients in 2019 did not have COVID, healthcare utilization in 2019 was higher than early on in the pandemic. And so these patients might have higher rates of health events, some of which may look like long COVID two are computable phenotypes than patients in 2021. So either way, these control groups are likely to overestimate long COVID compared to our COVID positive patients we identified during the same timeframe. Next slide.

For next steps, we are still revising this manuscript for resubmission and I wanted to mention that in response to our reviewer comments, we're actually working on a harmonized definition across our adult cohorts and are trying to align this as much as we can with the guidelines published by the National Academy of Sciences, Engineering and Medicine or NASEM report. So hopefully we'll have results of that supplementary analysis soon. And the last slide is just acknowledgement. So thanks so much to everybody who participated in conceptualizing and running the analysis for this manuscript, reviewing it, and everyone contributing data to recover. That's it for me.

Dr. Ravi Jhaveri:

All right, thanks so much to all of our speakers. Really interesting and provocative findings. I guess, I'll spill my beans, which is obviously that I'm in pediatrics. I'm the site PI for Lurie Children's in the PEDSnet EHR cohort. And I guess I'm always just struck by particularly that graph that you showed with

the incidents in children. And so much of the COVID discussion has really been about COVID overall being mild in children and we shouldn't worry about in children. I think that's motivated a lot of parents to not seek vaccination the way they perhaps should. And I'm really struck by the parallels in the pediatric data that you showed compared to adult data.

So one, I just wanted to highlight that finding because I think that's incredibly important for us in pediatrics, particularly with the patients and parents that we see who come in with symptoms of long COVID seeking guidance. Are you able to offer any more granularity about age? So many of the patients that we see who have persistent symptoms after other viral infections like mono or flu or other things tend to be adolescents. Are we able to dig down to see are those primarily adolescents or are we still seeing that same proportion of 4% even into the younger ages?

Dr. Richard Moffitt:

I don't think we have that prepared to answer you today, but -

Dr. Ravi Jhaveri:

Yeah, sorry, I didn't mean to put you on the spot.

Dr. Richard Moffitt:

... we can certainly the data enable us to answer that question relatively quickly at this point. I'm not aware that we've cut it up by age groups unless Hannah is remembering something that I'm forgetting. But yeah, we have the ability because this is all retrospective analysis to do sub-core analysis in adolescents versus younger kids.

Dr. Ravi Jhaveri:

Yeah, I'm sure you guys are well familiar that we struggle with the diagnosis of long COVID in younger children because they have symptoms that they don't easily align with sort of the definitions that you've put together. If they have more subtle school difficulties or behavioral challenges, it's hard to pull those out and align them with the definitions that are based on a older children, obviously on adults. And so that's constantly a struggle for us.

Dr. Richard Moffitt:

Yeah, and I would say the other key challenge with looking at the pediatric data is the kind of patient population that visits these children's hospitals is a little less of a well visit than the 30 or 40 year olds that visit adult hospitals.

Dr. Ravi Jhaveri:

Yes, for sure.

Dr. Richard Moffitt:

Yeah.

Dr. Ravi Jhaveri:

So let's see. Let's walk through some of the questions here so we can see. So one of the questions related to sort of medical complexity is what about the rates or likelihood of PASC among immune compromised patients? Can we pull those patients out from the groups you mentioned?

Hannah Mandel:

Yeah, for this manuscript, we didn't focus on that population, or yeah, the one I was speaking about. So I'm afraid I can't answer that off the cuff.

Dr. Ravi Jhaveri:

Understood.

Dr. Richard Moffitt:

Yeah, there's a current subgroup that's looking in particular in HIV positive patients, and there's other work throughout consortium which I'm less familiar with that is looking in other immunocompromised groups. I see one other question in the chat, which is basically a follow-up on your age question about older adults. And one thing that's interesting in these EHR-based studies, when you look at patients who are much older, like actually elderly or patients that have complex pre-existing conditions, sometimes we can start to see a paradoxical effect where things that we would call symptoms of long COVID stopped getting coded in the EHR due to a sort of...

There are other reasons that those things might be present in a patient's record, so they're less likely to be recorded. So if there's a patient, an elderly patient with diabetes that's uncontrolled and has very high A1C levels, it's much less likely that a physician is going to code fatigue in their chart, for example. So we see this sort of paradoxical effect where the sickest or the most vulnerable patients actually appear to have a little bit less long COVID as a result. And so we try and take care to avoid those whenever possible, but in general, whenever we slice up the patient population, unless otherwise noted, we see the general same effects in middle age versus older adults.

Dr. Ravi Jhaveri:

Thanks, Richard. I think there are several questions sort of along the theme of trying to account for home testing those with mild illnesses that didn't seek care. So rather than going through each one, can all of you discuss how you might account for those infections that occurred at home that didn't seek medical attention, but subsequently patients may present with for a visit for long COVID-like symptoms? How have you managed to adjust your analysis perhaps? Emily, do you want to go first and then we can go to Richard and Hannah?

Emily Hadley:

Yeah, sure. This was definitely a question that we struggled with as part of our analysis. And I'll note to one of the commentators questions. We didn't intentionally exclude people with home tests. We just don't get that information, people might test at home and not tell anyone what the result of their test is. And so, one of the pieces that we aim to focus on, particularly with our understanding of the severity of first infection versus severity of reinfection, was really to make sure we were especially focused on those groups who did have a more severe first infection because we considered them more likely to potentially have interactions with the healthcare system that might result in them documenting a future test, particularly because they might be engaging in a healthcare setting more often. And so you'll notice that we put asterisks on a lot of our findings where we're really not talking about population level incidence here.

We are really focused on people who have in general been interacting frequently with the healthcare system. We have also considered doing specific timing of our analyses. Earlier in the pandemic, we are more confident that we are capturing more positive tests than later. And so we've done some investigation into simulating what might've happened if we had more tests later in the pandemic. But I think that somebody else had mentioned the tests we're likely missing are most likely for people who have mild cases, which makes us more confident in some of our findings related to more severe outcomes and therefore less confident in some of our findings related to mild experiences.

Dr. Ravi Jhaveri:

Thanks, Emily.

Dr. Richard Moffitt:

Yeah, I think Emily has said it well. And one other sort of answer to this, which is also reflected in the question is we do have a limited amount of unstructured data from clinical notes. We don't have the free text notes. We have actually extracted pieces of information there, and we do see that, for example, with notes, we can reverse that paradox I talked about earlier where certain symptoms of long COVID are undercoded, but they're still there in the notes. So it might be noted that the patient had a cough somewhere in the notes, but that's not what gets coded for and billed. But in general, with all these things that EHR data are missing, our strategy is to focus on the questions that we can answer well and try not to do bad science on questions that we don't have business answering, for example, around home testing with EHR data.

Dr. Ravi Jhaveri:

Yeah, understood. There are several questions along the theme of both treatment and prevention. So I wonder if you guys could also address the issue about the impact of treatment for COVID symptoms and the subsequent risk of PASC, and then we'll move to the vaccine question after that. Do you guys want to talk about treatment?

Dr. Richard Moffitt:

Yeah, I think the best way to answer this is to say that those kinds of studies are underway or have recently concluded, I think might be the topics of upcoming R3 seminars in fact. But yeah, we have looked at different treatments and their effects on long COVID. And I would note, I think I answered in the chat here that EHR cohorts are a great way to sort of pre-test clinical trials. So the question was about use of off-label drugs and these data are a great way to see if we can see any sort of evidence of efficacy before we might go launch a trial. And also, it can help inform how that trial could be designed.

Dr. Ravi Jhaveri:

And then obviously the vaccine question is one that's out there. In a previous R3 seminar, we did cover some of the papers that are out there about the performance of vaccine in preventing symptoms of PASC subsequently. Again, those data are primarily from the early eras, but I wonder if you guys would comment on your work and then maybe I'd ask you to, well, maybe if you want to comment there and then maybe I have a follow-up question after that.

Hannah Mandel:

Yeah. So for the analysis I discussed, we definitely have caveats about the completeness of the vaccination data. I think around a quarter of adults had at least one COVID vaccination documented prior to their index event, and that number was lower for pediatric patients. So if somebody does have a documented vaccination, I think we're pretty confident they actually add a documented vaccination, whereas if a patient has an absence of documented vaccinations, we're really not sure what to conclude about that. So we were careful with the wording there and just considered, we didn't look at a number of doses type of vaccination or anything like that. In our adjusted analysis, we did find previously vaccinated adults had a slight but significantly lower risk of long COVID compared to unvaccinated patients. But yeah, I think given the limitations of our data, it's hard to make an conclusion.

Dr. Ravi Jhaveri:

Yeah. Emily, I'm going to put you on the spot a little bit and it's okay if you don't want to answer this question, but given that we now have updated boosters with the latest variants included, our vaccination rates really haven't moved at all. As you think about the reinfection data and subsequent exposure to newer variants, do you feel like we have some data to perhaps model how these vaccines would perform if we use them more readily and more people got boosters?

Emily Hadley:

Yeah, that's a great question. It's interesting you bring that up because we were actually just this past month doing a booster analysis as part of some separate query work for NIH and N3C. And this did allow us to see we have a fair amount of booster data in N3C. We haven't looked at the relationship of the boosters with reinfections and long COVID yet. I think that would be a future analysis, signed so far in our reinfections paper, again, it's not causal, I wouldn't say vaccination, but we did find that the percentage of people with reinfections had a lower proportion of people who are vaccinated than the percentage of people without reinfections.

So that would be suggesting that we might explore that further. And we've done some previous vaccination work that is being incorporated into our next round of causal reinfection analysis as to, I think we still have some of the EHR research limitations here though, right? Both related to who we get vaccination data for, who we get booster data for. We can't be confident that a lack of booster in the data means the person didn't get a booster. It might just be that they got a booster at a different facility that we didn't get that information for. So I put some pretty big asterisks on it. But we do have a fair amount of booster data in N3C, Richard, I don't know if you have more to add there.

Dr. Richard Moffitt:

Yeah, it's not about vaccination and boosters, but we do have evidence, if you remember back to the first figure, Emily showed that the frequency of reinfection, there is a sort of one or two-way wave lag protection effect that we see where the people that are getting reinfections are people that their previous infection was further ago, but that protection goes away relatively quickly. So we don't have the analysis to back it up quite yet, but one reason that there's a similar effect with vaccinations.

Dr. Ravi Jhaveri:

Yeah. Great. Thanks to all of you. So I think we're going to turn it back over to Quinn to lead us through the rest of the questions.

Quinn Barnette:

All right, thank you everyone for that rich discussion. I think we'd like to open up the rest of our audience Q&A with a few questions that we received in advance. And just a reminder to the audience, we will be posting a Q&A document on recovercovid.org for questions that we couldn't get to. Our first question is can be for anyone in the panel and asks, is the incidence of long COVID changing as we get further into the pandemic based off of your studies or just in observations even outside of your studies? And if so, could this be related to vaccine uptake and/or repeated infections?

Dr. Richard Moffitt:

You want to take a first shot at this Hannah or you want me to?

Hannah Mandel:

Yeah. So yeah, I think there's a lot of temporal factors that could be involved in what we're seeing. So I think I mentioned some of them at the end, definitely that could be influencing things. We did not see a sizable decrease in long COVID when we're looking at those secular trends over time. So although those could be playing a role, maybe what we're seeing is not related to that. It could be, again, increasing recognition and documentation [inaudible 00:50:46]. And although we're still not seeing [inaudible 00:50:55]. You can't hear me?

Dr. Richard Moffitt:

Oh, you're back now.

Hannah Mandel:

Yeah. Yeah. So I mean, there's increase of healthcare utilization, increasing awareness, multiple infections over time, which again, we did not look at for this, but there's a number of factors.

Dr. Richard Moffitt:

And sort of another way to ask that same question is does the subsequent reinfection give you another or boosted chance at getting long COVID? And that is the follow-up work that we are currently working on based off of Emily's work that we've shown today. But the simple answer is that every reinfection is a new encounter which gives a non-zero chance to experience long COVID.

Quinn Barnette:

Right. All right, thank you. Our next question is for Emily, and I think you answered this a little bit in discussing home testing and the limitations there, but this person asks, that one concern is that something being labeled as first infection can actually just be the first documented infection, and so this could undermine the finding of lower or long COVID prevalence in the reinfected group. Have you considered this limitation at all in your paper?

Emily Hadley:

Yeah, this is a great question. It was less of a concern with some of our initial paper just because early in the pandemic, there were not that many home tests available. And so we have more confidence in many of the tests that we have during the ancestral period and the Alpha period and some of the early Delta period as well. It's definitely become more of an issue as we've gotten later in the pandemic, both from a consideration of, we don't know how many people still have yet to have experienced a COVID infection. And so when we do have an initial infection that pops up much further out than the initial pandemic period, we are curious if we might've missed some initial infections. It's a hard limitation to overcome.

I think that Richard pointed out earlier, it's possible we might be able to glean some information from the notes where a clinician wrote, this is actually the third time this person's had this infection, even if it's the first time it's documented in their record. But even that might be somewhat unlikely. So it is a limitation of our research, especially as we move further out from the initial date of the pandemic. And again, kind of goes back to that piece of we're very cautious about how we generalize our findings and are really focused on people who are likely to interact with the healthcare setting and have a rather robust record.

Quinn Barnette:

Great, thank you. Our next question is directed for Emily and Richard and asked, are there any potential explanations for why we're seeing the average age of reinfection is younger than the initial infection?

Emily Hadley:

We had a couple of different suggestions for this in our paper, and I can take this first. One of them was related to behavior. We thought that younger folks may be more likely to engage in behaviors like going to work or engaging in daily life settings that perhaps other individuals might be differing to other people. There was some evidence that suggested early on in the pandemic that younger folks may be less likely to wear masks or take part in some of the other recommendations to prevent the spread of COVID or might just be less likely to be concerned about getting COVID again. And so might not be taking the same precautions that generally folks who were older in age were taking. So that was some of our thoughts around likely increased risk of exposure to second infections and then considerations too related to vaccination that we covered earlier. Richard, other pieces?

Dr. Richard Moffitt:

Yeah, I think the only thing that maybe you didn't touch on was a lot of folks in the working population age are surveilled as part of their jobs or were surveilled as part of their jobs. So there might just be better selection for folks that have more testing available even for asymptomatic testing.

Quinn Barnette:

Great, thank you. I'm going to switch. Our next question will be for Hannah and Richard and ask how might the increased recognition of long COVID over time impact the reliability of longitudinal trends observed in the data?

Dr. Richard Moffitt:

I would say that the increase over time, we basically have a monotonic increase in information and all of the computable phenotypes and machine learning approaches that Hannah mentioned are at our core based on this long COVID diagnosis. And so the more that that happens, the better for us data analyst types. And yeah, I think one thing that's going to be perhaps different over time is maybe the type of people that are going to get those diagnoses might change. We do have other work from ourselves and colleagues that have taken a deeper dive into looking at the difference between the population that actually gets a diagnosis versus the general population of COVID patients, also versus the population that the models suggest could have a long COVID diagnosis but don't have it documented.

Quinn Barnette:

All right. Thank you. The next question is also for Hannah and Richard, are you aware of other evidence that confirms your findings related to lower albumin levels surrounding COVID infections? And do you know if this is of interest to trialists as their potential therapeutic target?

Dr. Richard Moffitt:

I might say that I think there's a difference between a significant difference and a clinically significant difference and the differences that we see are statistically different, but it's unclear that the magnitude of the change would be clinically relevant, and in fact, it might take a whole handful of these several biomarkers. And if you combine them all together, that might be enough information to start to look at a risk factor biomarker panel or even a diagnostic factor biomarker panel. But albumin alone is never going to be by itself important enough to really make that decision. It's going to have to be one as part of many pieces of evidence.

Quinn Barnette:

Great, thank you. Our next question is for Emily and Richard, and it asks, do you adjust for the treatment type received in your adjusted analysis for long COVID?"

Emily Hadley:

Yeah, I should clarify that this is actually not an adjusted analysis. This is just an observation of how frequently we are seeing long COVID diagnosed following an initial infection versus reinfection. Our current analysis that we are working on publishing will include a series of adjustments and we'll have more of a causal interpretation, but this was really to understand do we see long COVID diagnosed among people with brain infections in our data? And it actually doesn't even include the computable phenotype that Hannah discussed. And so we'll be including that in the next round of analysis as well. But I would say this was very much meant as a preliminary can we do research related to reinfections and long COVID with this work?

Quinn Barnette:

All right, thank you. Our next question I think is also for Emily and Richard and asks, how does this research incorporate findings that PCR testing can be inaccurate in identifying COVID cases in children and how might that impact the study's results?

Dr. Richard Moffitt:

Yeah, it's always a good question to consider how our inclusion criteria are working, conspiring to help us or hinder us in certain conclusions. And I think probably the best answer again is that we really try and focus on the pieces of the analysis that we feel are good and sound. And if we're underestimating diagnoses in a certain population based on the testing, unless we have solid evidence that this is a non-random phenomenon, we hope that and generally it washes out in the large numbers.

Presumably there would be some relationship to the negative tests also being related to less severe disease perhaps related to differences in viral load. So in this way, it would just lead us back to the same sort of bias about not being able to capture home tests or mild symptoms. Hannah?

Hannah Mandel:

Just to add on to that, we tried to also include a number of different inclusion criteria. So not just testing. We looked at diagnoses and then for children, so that would include acute COVID and also long COVID as well as MIS-C. And we also looked at igG test results. So if patients had indication of one of those, we thought they must have been COVID positive at some point. And if they had no other indication of an acute infection, we tried to impute the date of their acute COVID infection for the purpose of this analysis. So yeah, that's in more detail in the pre-print, but we tried to cast a broader net.

Dr. Richard Moffitt:

Whereas in contrast in the reinfection work, we were much more strict in terms of defining when that reinfection occurred because we wanted to be very aware of the fact that sometimes diagnoses are carried forward in EHR. So just seeing the COVID diagnosis, it's not exactly clear when doing the analysis if that was a new COVID diagnosis or mentioning that COVID happened sometime before. So we really leaned heavily on the actual test results for determining reinfection.

Dr. Ravi Jhaveri:

Yeah, if I could just jump in for just a sec. I think there are so many variables too that go into the test that have nothing to do with the performance of the test. You obviously have to have a high quality sample, and for young kids who are screaming and crying, an adequate sample sometimes is a real challenge. And so that in and of itself can affect test performance. And similarly, we have many challenges with patients who persistently test PCR positive, and the result, it doesn't really impact or isn't relevant to the current clinical status that's going on. So I think we are left with making our best guess on both sides of the equation. And I think those caveats come into play whenever we're talking about a test, in particular, COVID PCR testing.

Quinn Barnette:

All right. Thank you all. Our next question is perhaps slightly out of scope of the single two papers that you presented, but I think it's interesting and perhaps I'll throw it out there and see what you all think. So this is for anyone on the panel and it asks, do you think that the results from your studies could be used to improve documentation for medical professionals with patients who have long COVID? They state that in their experience documentation is very lacking and it affects the ability to seek specialize care or receive disability benefits.

Dr. Richard Moffitt:

Yeah, I hope so. I think maybe not these two studies that we talked about, but other work from our groups and our colleagues. If you look just back at the beginning, the computable phenotype

definitions where we're really just trying to find evidence of long COVID in EHR, all of those come with lists of diagnoses which are associated with long COVID. And I think what Hannah presented at the very end to the new definition from NASEM that's coming out is extremely inclusive and echoes a lot of the findings that our groups have found empirically from the EHR.

Dr. Ravi Jhaveri:

Yeah, I would chime in with the same hope that yes, that's part of the reason why we're in this doing this work is to try to help inform not only for patients but also for providers about the spectrum of PASC and to certainly improve it. I think the insights that we've also had are certainly the hope from some of our natural language processing data to show that when you actually look at some of the notes, there may be many mentions of long COVID, but in fact there's no code that's used when a patient is seen and the like. So there's a lot of discrepancies that we think by shedding light on and doing the research to show these patterns that will help improve provider recognition and subsequent documentation that enables patients to get better care and the support services that they need. So it's definitely very much the goal for all of us.

Quinn Barnette:

All right. Thank you. We have one more question that is slightly out of the initial two papers, but again, I think it's an interesting and important question, and I think perhaps this one would go to Richard and it asks, are you aware of any evidence that repeat infections may prolong or increase the effects of long COVID?

Dr. Richard Moffitt:

So the duration and severity of long COVID is unfortunately another thing that's really tricky to study in EHR. So first of all, there's only that one diagnostic code. Either you have it or you don't. So there's no severity encoded within that. But more than that, the first crush that people generally go to go to with trying to decide if long COVID is resolved or not is looking for an absence of new complications that define long COVID. And we want to be really careful to avoid conflating stopping seeking care with that person's long COVID having resolved. And from an EHR perspective, we have really no way to disambiguate between someone who's fed up and stopping seeking care versus someone who's continuing to seek care and generating more data in the EHR. So this one falls again under the category of things we are going to avoid studying directly for now in favor of things that the data are better suited to answer for us.

Quinn Barnette:

All right, thank you. To wrap up, we have one more question, and perhaps, I'll direct this to Richard, but I welcome any panelists to speak up and it is based on these data, what recommendations might you make to healthcare providers for monitoring and diagnosing long COVID moving forward?

Dr. Richard Moffitt:

I would say one of the key summary from the work Hannah presented is that long COVID is still happening at a rate that is not drastically reducing from where it was at the beginning of the pandemic with respect to the fraction of patients who are getting it versus getting infected. And at the same time, from the work Emily presented, every time you get a reinfection, you have yet another chance to get

long COVID. So both of these details conspire to suggest that avoiding infection and subsequent reinfection is your best way to prevent getting long COVID.

Emily Hadley:

And I would just add, I think a lot of our work, when we were first putting this out, there were still a lot of questions around do people get reinfections? How often do they get reinfections? What is the reinfection experience like? And I think our paper is really helpful evidence that people do get reinfections and they get reinfections more than once and that the experience might change with different variant processes and that you can still to, Richard's point, be diagnosed with long COVID afterwards, but you can also still be hospitalized. You can also still have a really negative experience with this illness. And so I think that people continue to move past this, just like I've got COVID once, I'll never get it again. That's not what we're seeing and so I think it is really worth paying attention to reinfections and educating clinicians and others that COVID is still very much out there and people still get it.

Dr. Ravi Jhaveri:

Yeah, I would echo all of those comments and say this, long COVID is actually a big part of the discussion I have with patients and families encouraging vaccinations or boosters as well. Again, the data that's out there suggests that there is a positive effect. It's not a 100%, we know that, but still as both Richard and Emily and others have said, whatever impact we can have to reduce your risk overall, it's certainly worth considering.

Quinn Barnette:

All right. Well, thank you all so much and thank you to our audience for attending this seminar and engaging with the Q&A. As a reminder, we'll be posting a Q&A document that has responses to questions that we received, and we'll also include some that we did not have time to address today. Before we conclude, a reminder that researchers both within and beyond the recovery initiative can now apply to use Recover data for ancillary studies. This includes data from three Recover cohort studies adults including pregnant adults, pediatric, and autopsy and biospecimens collected from cohort study participants. Interested researchers must submit an ancillary study proposal and receive approval.

Researchers must also have independent funding support to conduct the proposed study, and to learn more about that, you can visit recovercovid.org/ancillary. Our next slide lists future sessions, we'll have some exciting topics coming up and we hope to see everyone there for future sessions. And here shortly, you'll see a short survey coming up on your screen which asks for your feedback on this seminar and we would really appreciate you take a minute, fill it out. And with that, thank you again to all of our panelists for this really amazing discussion and presentations. And thank you to our audience for tuning in. All right, have a great day everyone.