

# Responses to Participants' Questions

The overarching goal of the R3 Seminars is to catalyze a shared understanding of the research being conducted by the scientific stakeholder community within the RECOVER Consortium. The R3 Seminars and the Q&As typically feature highly scientific material intended for researchers and clinicians. For other audiences interested in these topics, a link to the National Library of Medicine's MedlinePlus medical dictionary is provided at the end of the Q&As as a resource to help in understanding the scientific terminology.

This document provides responses\* to questions raised by seminar participants related to the following presentations at the R3 Seminar *Patterns and Prevention of Long COVID: Findings from RECOVER EHR Cohort Studies* held on March 19, 2024:

- ***Association Between SARS-CoV-2 Infection and Select Symptoms and Conditions 31 to 150 Days After Testing Among Children and Adults***  
Yongkang Zhang, PhD
- ***COVID-19 Vaccine Effectiveness Against Long COVID in Children***  
Hanieh Razzaghi, PhD, MPH
- **Discussant: Ravi Jhaveri, MD, FPIDS, FAAP**

\* Responses may have been edited for clarity.

## All Presenters: Questions and Responses

**Q. How are you accounting for variant emergence and the changes in symptoms as you carry these studies forward?**

**Response:**

**Dr. Zhang:** Yes, that's a very good question. With electronic health record (EHR) data, it's very difficult to determine what variant this person had because we only know if the test result is positive or negative, but there's no additional information such as Omicron or Alpha or Delta. So, to control for when this person tested positive, we use a time stamp—like March 2020, April 2020, January 2021. This could be a proxy for the variant, but it's not perfect because EHR data does not have such information available for researchers.

**Dr. Razzaghi:** Yes, so we've attempted to look at this by understanding or trying to encompass a wide range of symptoms because of the changing variants. The vaccines didn't come until well into the pandemic. So, especially in the 5-to-11 age group, there has been a lot of multisystem inflammatory syndrome in children (MIS-C), for example, that wouldn't be that much of an interest—and because of that, it has been excluded from our definition. So, we have tried to align the time periods with the disease clusters that we came up with to account

for the differences in presentation that we've seen. And then the Long COVID diagnosis code obviously is underdiagnosed—it's not used as frequently as it should, but it does offer the added benefit of a clinician diagnosing and saying, "This patient has Long COVID."

### **Q. Can you discuss how you tried to account for the patient reporting of at home testing?**

#### **Response:**

**Dr. Zhang:** This is such an important question. I think all the EHR-based studies have this question, "How do you know if this person, which you identified in the EHR, is truly negative?" when you consider capacity constraints we had in the beginning of the pandemic, anywhere in the country. We did a few things to account for this potential bias. First, we did a sensitivity analysis by excluding patients who were tested in an early phase, like the first three months—March, April, May in 2020—because we know during this time period, testing capacity was very limited and there was no standard of how to code a positive test back in those days. The physicians were using a variety of ICD-10 codes to indicate that this person has COVID.

So, we saw that the testing results and diagnoses in the early phase of the pandemic were less reliable compared to the later phase. For many of our studies, we excluded the first three months of the pandemic in 2020, to ensure our results are consistent. Also, to make sure a positive is a true positive and a negative is a true negative, among those who tested negative, we made sure that those people didn't have a diagnosis suspected of being COVID. We also made sure they didn't receive a therapeutic for COVID, such as remdesivir or some other treatment the physicians used in the early phase of the pandemic to treat COVID.

We did a lot of sensitivity analyses to see if the results were inconsistent. Again, this issue is very hard to address, but we were able to do all kinds of analysis to make sure we did our best to address it.

**Dr. Razzaghi:** For our study, we required the symptom-based diagnosis for Long COVID to have any kind of evidence of COVID-19. That could be a lab result or just a diagnosis code that the clinician entered. So, if a patient was seen for COVID-19, regardless of testing or where or how, it would have been recorded. So, that was our attempt to do that. Of course, we're not accounting for the patients who had COVID-19 but didn't report it or weren't sick enough to see a clinician.

However, we did exclude the patients who had a history. If they were being seen repeatedly for other things and they had a history of COVID-19, we did exclude those patients. We didn't require original infection for the diagnosed Long COVID group.

**Q. How did Dr. Zhang disentangle Long COVID from prior conditions and patients with medical complexity in the study?****Response:**

**Dr. Zhang:** For condition outcomes like mental health, diabetes, and chronic kidney disease, we're looking at newly diagnosed conditions. We made sure there was no diagnosis for this condition prior to COVID testing. In the baseline, 1-month to 18-month period, we made sure this person did not have any diagnosis for diabetes, and if we observed a newly diagnosed condition, and if incidence of newly diagnosed diabetes is much higher among positive versus negative, it's more likely to be a PASC condition.

I think it's hard to disentangle for symptom conditions because fatigue and shortness of breath are very common symptoms. It could happen to anyone for any condition, and EHR data probably does a poor job of capturing symptoms compared to capturing conditions because many physicians probably don't bother to code the data in EHR. So, symptoms are probably harder to disentangle or ascertain compared to conditions.

**Q. How do you account for patients who have underlying conditions not found in the EHR that are exacerbated post-COVID?****Response:**

**Dr. Zhang:** It has to be defined as a newly diagnosed condition or exacerbation of preexisting conditions. So, we started to look into this question. We started with diabetes, for example. We looked at people who had type II diabetes before the pandemic and defined a few indicators that may potentially indicate exacerbation of type II diabetes, such as elevated HBA1C testing results, or increased use of anti-diabetic medication. An example would be only one medication class becoming two or more after a positive COVID-19 test. We started with diabetes because diabetes outcomes are easier to define based on EHR data because you have lab testing results and medication prescription in EHRs. For depression and anxiety, for example, I think that will require some questionnaire-based assessment results.

So, it's probably a bit harder to do with EHR data alone. We will need to combine EHR with some other data. Of course, we can look at the prescription data. For people with mental health conditions, we can look to see if there are any changes in terms of medication use before and after COVID test data. So, that could be something we wish to explore in the future. Again, this is a very important question. Not only newly diagnosed conditions, but also the exacerbation of prior conditions after COVID.

**Dr. Razzaghi:** This is a very difficult question around disease exacerbation. So, there are a couple of ways to think about it. We've been doing disease-focused analysis. We did one on type I diabetes, for example, showing how the A1C does actually get a little bit worse, but then normalizes over time for most patients after COVID-19. We are currently doing one on sickle cell. There are certain kinds of conditions that we're doing deep dives on to look at

this for these large-scale analyses. We have a washout period which varies based on the timeline of a patient's healthcare visits. So if we don't see medications completely new to the patient, but for example, if they weren't seen for 6 months for their depression because they were handling it okay, but then we see a new utilization, we will say that it still counts. It's likely potentially an exacerbation. So, that's the way that we attempted to balance the question about the exacerbation.

**Q. Please discuss successes and barriers to obtaining info from EHRs for research and comment on strategies for inter-institutional data sharing.**

**Response:**

**Dr. Zhang:** This is definitely a challenge. I mean, if you consider New York City, we have such a fragmented market. We have so many big hospitals in New York City. Cornell, Mount Sinai, NYU, Columbia, Montefiore are five major ones in Manhattan, not to mention Long Island. So, people could go anywhere to receive care. For researchers, we would like to capture the comprehensive information as much as possible to track patients across different institutions to have whole-person information. That means that only using one hospital's data is not complete. You miss lots of information about positive tests, medication use, diagnoses.

So, we leverage the PCORnet framework. If hospitals are members of this network, they are able to track the same patient across different hospitals. Using this framework, we have much better, much more comprehensive information for each person in our data. Of course, many people could also utilize healthcare in some other hospitals that are not members of this network. In those cases, we will not be able to capture information outside the network. Something else we've been considering is using claims data, which will be much more comprehensive. So, I don't think there's an answer for this question. I think for each project, we just identify the key data elements we need for each research question and do our best to combine data from different sources to find the most reliable, most robust data for the question.

**Dr. Razzaghi:** EHR data is very challenging. We can't get around it, right? The EHR data creates what we call collider bias in terms of who gets healthcare, and therefore, we're already starting with bias. Given that, we try to really account for that in the analyses that we do. So, we make sure that the two cohorts where we're looking for differences in effect are similar enough or are comparable so that we can draw on observations from the data that we have. So, there's the utilization bias. There's what gets recorded.

In addition to accounting for these statistically, we try to do what we call study-specific data quality analysis to understand the extent of the bias that we're dealing with or how diagnosis codes are represented or not represented across different institutions. Are they using things a little bit differently to represent the same idea? So, we really tried to mitigate these kinds of biases by spending a lot of time upfront evaluating the different kinds of biases and what we can do to address them.

**Q. With the burgeoning number of research studies being done on Long COVID, would it be helpful for the research community to communicate to clinicians to code for Long COVID in EHR?****Response:**

**Dr. Razzaghi:** Yes, and that's a really great observation, because you can only study something if it's recorded. So, that's why we tried to come up with these data mining techniques to understand what is occurring more frequently in these groups and then generalize it to a larger population so that we can overcome these barriers. I think that clinician outreach has to be an important part of what we do and within our health systems, we need to feed back this learning health system model where we provide information back to clinicians.

I bet if you look at clinician notes, they probably mentioned Long COVID. So, as part of the RECOVER program, what we're doing now is going through and extracting observations about patients from clinician's notes that we can't get just from the structured data. So, we're trying to get inside the mind of the clinician, whether or not they are diagnosed officially in the EHR.

**Q. Why are anxiety and depression not included in the probable diagnosis disease cluster slide?****Response:**

**Dr. Razzaghi:** It's very important to address this question about the mental health impact. There was a cluster for affective cognitive functioning. As mental health impacts functioning, we're hoping that those were captured, because they were occurring more frequently. When we do these analyses, what we've also noticed outside of kids who were diagnosed with COVID-19 was the large rise of mental health issues, just from the social impact of the pandemic in general. So, this was very difficult to disentangle from the direct impacts of COVID.

A lot of these signals actually were not prominent when we were comparing across the groups. However, that doesn't mean that it's not real. So, as I mentioned in my previous question about the notes, and the specific disease or condition-focused exacerbation is that we are trying to make mental health a priority, given its impact and the way that the social impacts of the pandemic have also made it difficult to disentangle. A deeper dive is warranted into this. We're doing that and also with the clinician notes, using these mental health features as well as extracted data from clinical notes, helps us get at what's going on a little bit better. So, that is a limitation of the way that we were forced to demarcate these and label these probable conditions.

**Dr. Jhaveri:** We've spent this session talking about EHR studies and some of the advantages, but also some of the disadvantages. It's important to remember that RECOVER is many things, as I talked about at the beginning, and one of the specific advantages of some of the prospective studies is to be able to ask some of these questions in a more detailed way or in a different way to try to account for some of the shortcomings of relying exclusively on

EHR-based research or on lab-based research. So, thinking about the comprehensive nature of what the RECOVER Initiative is trying to do is really important and the whole initiative is trying to come at it from different directions.

**Q. When controlling for mechanical ventilation (MV), did you also test for the effect of the number of MV days on COVID diagnosis?**

**Response:**

**Dr. Zhang:** Yes, that's a very good question. I think the tricky part is that in the EHR data, we only observe who had the mechanical ventilation based on the procedure code, a CPT code, or the ICD-9 or ICD-10 procedure code for MV. It's very hard for us to count how many days this person used MV because I don't think EHR data has a per-day record. We only know what happened during the encounter without knowing, "Okay, probably this procedure was repeated for 5 days." It's a really important thing to consider.

**Q. Have symptoms of Long COVID been observed to worsen or reemerge following a non-COVID infection or due to other triggering factors? If so, what are these triggers?**

**Response:**

**Dr. Zhang:** We have to look at into this, but if you think about it like any other hardship, everyone experienced the pandemic of social isolation, for example, regardless of our COVID status—especially during the first year of the pandemic. It could have exacerbated everyone's mental health status regardless if they were COVID positive or negative. There were other hardships as well, like unemployment experienced in 2020, lack of access to food, and lack of access to healthcare for elective purpose. So, all this hardship we experienced could exacerbate a lot of conditions, even for people who were COVID-negative at that time.

**Dr. Razzaghi:** We haven't done a lot of research into this. One of the things that we're looking at is the effect of subsequent infections after initial COVID-19 infection. For example, if you look at children with COVID-19 and another respiratory illness, both the frequency and the severity and the types of illness, for example, respiratory syncytial virus (RSV), is it more severe after initial infection? We're actually currently doing this analysis and I don't have any preliminary results yet to show, but this is something that as the pandemic goes on, we are trying to understand how COVID-19 is or is not different than other respiratory infections.

So, this is going to be an important question about what triggers Long COVID, what triggers vulnerability to subsequent infections, and things like that.

**Q. Do we know the acute and long-term effects of administering mRNA and non-mRNA vaccine on patients who are presently with Long COVID? Is there an exacerbation of symptoms?****Response:**

**Dr. Razzaghi:** So, what we have observed is that if you get a vaccine after initial infection, you are protected from Long COVID—but I think the question is asking if you currently have Long COVID, what will happen if you get the vaccine, and we have not yet studied that. So, I don't think we have any data to back that up. I will say that anecdotally, there have been reports that there's been improvement of symptoms after the administration of the vaccine, but we have not specifically looked at this question.

**Dr. Jhaveri:** I would also just add, I think that there are few individuals who have gotten the protein vaccine. The submitter of the question is asking specifically about it and it's really hard to denote any specific effects of that. So, obviously, we have millions upon millions of people who have gotten the mRNA vaccine, and many, many, many fewer have gotten the Novavax protein vaccine. So, it's really hard to know. Other than the general protection level that it offers, we don't really have the data yet for the downstream benefits.

**Q. What does RECOVER plan to do about COVID-induced PTSD (from COVID infection or Long COVID)? Is PTSD being tracked or studied?****Response:**

**Dr. Zhang:** I agree that PTSD is an important condition to consider, and we have ongoing research focusing on mental health conditions. We will pass this comment to the team!

**Q. Would a March 2020 false negative test exclude that patient's records?****Response:**

**Dr. Zhang:** We conducted sensitivity analyses by excluding data from the first few months in 2020 and the results were largely consistent.

**Q. Were any of the patients vaccinated and/or boosted with the mRNA products?****Response:**

**Dr. Zhang:** We were not able to adjust for vaccine status as EHR data poorly captures vaccine status. We are working on linking EHR data from NYC to vaccine registry so we can adjust it in the future.

**Q. In non-hospitalized patients, is the non-recognition of severity of symptoms and attributing them to "anxiety" a factor that can confound results?**

**Response:**

**Dr. Zhang:** I agree that EHR data is not good enough to identify the severity of symptoms as we only have ICD codes. Future studies should consider incorporating clinician notes to identify severity.

**Q. Could differentiation between hospitalized and non-hospitalized patients have bias regarding inferred severity? I recall many patients with severe respiratory and vascular symptoms refusing hospitalization due to congested hospital censuses.**

**Response:**

**Dr. Zhang:** Thank you for this important comment. This is definitely possible, especially in the early phase of the pandemic (2020). We conducted sensitivity analysis by excluding patients who were tested in the first few months in 2020 to mitigate this bias.

**Q. Tracking post COVID with "fatigue" may underestimate the issues of muscle loss, strength loss, and balance/uncoordinated movement. Is there further work to define and track these post-COVID effects?**

**Response:**

**Dr. Zhang:** Thank you for this important comment. I totally agree that EHR is not perfect to track symptoms such as fatigue. We are exploring other methods and data sources (e.g., notes) to better identify fatigue and other symptom conditions.

**Q. Blue line for "all" was less effective than blue lines for each age category. Why?**

**Response:**

**Dr. Zhang:** The blue line and the red line for "all" was between the 5–11 and 12–17 year-olds. For the 18-month outcome, it was not, but the confidence intervals overlap. We used different institutions for the 5–11 and 12–17 groups, so it's not a direct comparison of the exact same patients, which may drive small differences.

**Q. Dr. Zhang, did you take into account patients who received COVID-19 treatment? I believe your study included a timeframe when monoclonal antibody therapy was probably available. Also, early use of remdesivir might have been available.**



**Response:**

**Dr. Zhang:** We controlled for treatment (medication and ventilation) for hospitalized patients. But I agree with you that monoclonal antibody therapy and other treatments should also be included in future analyses.

**Q. Do the EHR data to which you have access contain results of complete blood count (CBC) blood tests, such as red cell distribution width, mean corpuscle volume, and mean corpuscular hemoglobin concentration? And if you reviewed them, do they show any abnormal patterns?**

**Response:**

**Dr. Zhang:** This is a very interesting question. We just did a query to look at abnormal results of other labs. We should have this published very soon. We did see some abnormal patterns.

**Q. Dr. Zhang, when controlling for mechanical ventilation, did you also test for the effect of the number of MV days on COVID diagnosis? I am thinking this should have been significant.**

**Response:**

**Dr. Zhang:** Thank you for this comment. I don't think we controlled for MV days. I will pass this important comment to the team. Thank you!

**Q. Have you studied if there is a common experience of someone with Long COVID who then gets a booster, and each time then has a difficult reaction—exacerbating symptoms for a few days—each time? So, could the vaccine somehow be a problem for a person with Long COVID?**

**Response:**

**Dr. Razzaghi:** We have not observed this, but we have not studied it in detail yet.

**Q. Have you seen an increase in kids having ulcerative colitis or Crohn's disease because of the virus?**

**Response:**

**Dr. Razzaghi:** We have seen GI symptoms consistently emerging as a subtype of Long COVID in children.

**Q. At any point during the study, did anyone actually review some percentage of the medical records to make sure that the coding was truly reflective of the medical visit? Essentially, auditing the data.**

**Response:**

**Dr. Zhang:** We are doing chart review for certain conditions, such as diabetes.

**Q. Have symptoms of Long COVID been observed to worsen or reemerge following a non-COVID infection or due to other triggering factors? If so, what are these triggers?**

**Response:**

**Dr. Zhang:** We have not looked into this, but I think it could happen due to social isolation or other social hardships non-infected people have experienced.

**Q. What are other data sources outside of EHR data did you consider before embarking on these studies?**

**Response:**



**Dr. Zhang:** We also considered social determinants of health data and used them to examine socioeconomic status disparities. We are considering claims as well.

**Q. Duration of symptoms and whether they resolve by themselves or after treatment is an area that needs further research. What is the plan for this? It is important to inform public health policies. It is a general question, maybe not specific to these two studies.**

**Response:**

**Dr. Razzaghi:** Duration is one of the priorities for future studies, particularly in children, as we hope to ascertain functional status from clinician notes.

#### **To Learn More**

- Information about RECOVER research and to volunteer for studies: <https://recovercovid.org/research> 
- Frequently Asked Questions about RECOVER and PASC: <https://recovercovid.org/faqs> 
- CDC information: Information for the general public and for healthcare providers about Post-COVID Conditions: <https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/>
- For medical/scientific terminology: <https://medlineplus.gov/healthtopics.html>