Seroprevalence of prior infections of common human coronaviruses in children in Southwestern Pennsylvania and the associations between race, sex and age on seropositivity

by

Lawrence Ji

Bachelor of Science, University of Pittsburgh, 2021

Submitted to the Graduate Faculty of the

School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Public Health

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Lawrence Ji

It was defended on

May 23, 2022

and approved by

Mackey Friedman, PhD, Assistant Professor, Infectious Diseases and Microbiology, University of Pittsburgh, School of Public Health

Ernesto Marques, MD, PhD, Associate Professor, Infectious Diseases and Microbiology, University of Pittsburgh, School of Public Health

Ken Ho, MD, Assistant Professor, Department of Medicine, University of Pittsburgh, School of Medicine

Thesis Advisor: Mackey Friedman, PhD, Assistant Professor, Infectious Diseases and Microbiology, University of Pittsburgh, School of Public Health

Seroprevalence of prior infections of common human coronaviruses in children in Southwestern Pennsylvania and the associations between race, sex and age on seropositivity

Lawrence Ji, MPH

University of Pittsburgh, 2022

OC43, HKU1, NL63, and 229E are human coronaviruses (HCoVs) endemic to the United States and can cause mild and sometimes severe disease. This project investigated seroprevalence of these four coronaviruses in children in Southwestern Pennsylvania and explored associations between race, age, and sex on seropositivity.

Residual blood samples were collected from children under the age of 16 who visited the University of Pittsburgh Medical Center Children's Hospital of Pittsburgh for routine clinical care. ELISA assays were run on samples to determine optical density (OD) levels. The OD cutoff for positive and negative cases were determined by observing where the divergence between high optical density and low optical density lies. Using this cutoff, the prevalence of prior infection was determined for each virus.

The prevalence of HCoVs between April 27 and July 3, 2020, was 79.46% for HKU1, 61.61% for 229E, 80.36% for OC43, and 83.93% for NL63. For every one year increase in age patients are 35% less likely to have antibodies consistent with prior infection for NL63, 30% more likely to have antibodies consistent with prior infection for 229E, and 18% less likely to have antibodies consistent with prior infection for HKU1. For every year increase in patient age there is a 0.07 increase in optical density when detecting for IgG against 229E spike proteins and a 0.06 increase when detecting for IgG against NL63 spike proteins. The results demonstrate children are more likely to be positive for NL63 and 229E as they age. There was no significant association

between race and sex on seropositivity and optical density. The findings of this project can inform health practitioners and public health officials on the circulation of common human coronaviruses in the region. Common human coronaviruses cause more severe symptoms in children and thus it is useful to surveil the epidemiology of coronaviruses in these groups.

Table of Contents

1.0 Background 1

2.0 Methods 7

2.1 Statistical Analysis 9

3.0 Results 11

3.1 Prevalence of Prior Infection 13

3.2 Effect of Variables on Seropositive 16

4.0 Discussion 21

Appendix SOP: Total IgG spike and nucleocapsid of common cold coronaviruses-indirect ELISA 25

Bibliography 29

List of Tables

Table 1. Sociodemographic data of 112 patients less than 16 years old in Southwestern Pennsylvania, April-July 2020 8

Table 2. R squared values for the correlation of optical densities of ELISAs..... 12

Table 3. Association between age, race, and sex on HKU1 seropositivity 17

Table 4. Association between age, race, and sex on 229E seropositivity 18

Table 5. Association between age, race, and sex on NL63 seropositivity..... 18

Table 6. Association between age, race, and sex on OC43 seropositivity 18

Table 7. Results of the multiple linear regression with race, sex, and age as independent variables and optical density measuring IgG against 229E as the dependent variable 19

Table 8. Results of the multiple linear regression with race, sex, and age as independent variables and optical density measuring IgG against NL63 as the dependent variable 19

Table 9. Results of the multiple linear regression with race, sex, and age as independent variables and optical density measuring IgG against HKU1 as the dependent variable 20

Table 10. Results of the multiple linear regression with race, sex, and age as independent variables and optical density measuring IgG against OC43 as the dependent variable 20

Appendix Table 1. Recombinant Proteins 25

List of Figures

Figure 1. Weekly cases of HKU1, 229E, OC43, and NL63 coronaviruses in Pennsylvania, 2014-2022 2

Figure 2. Pearson's R correlation comparing assays by optical density readings 11

Figure 3. Representation of patients by age and optical density result 13

Figure 4. Seroprevalence of HKU1, 229E, OC43, and NL63 coronaviruses in children 1-15 years old in Southwestern Pennsylvania, April-July 2020 (n=112) 14

Figure 5. Seroprevalence by age of HKU1, 229E, OC43, and NL63 coronaviruses in children 1-15 years old in Southwestern Pennsylvania, April-July 2020 (n=112) 15

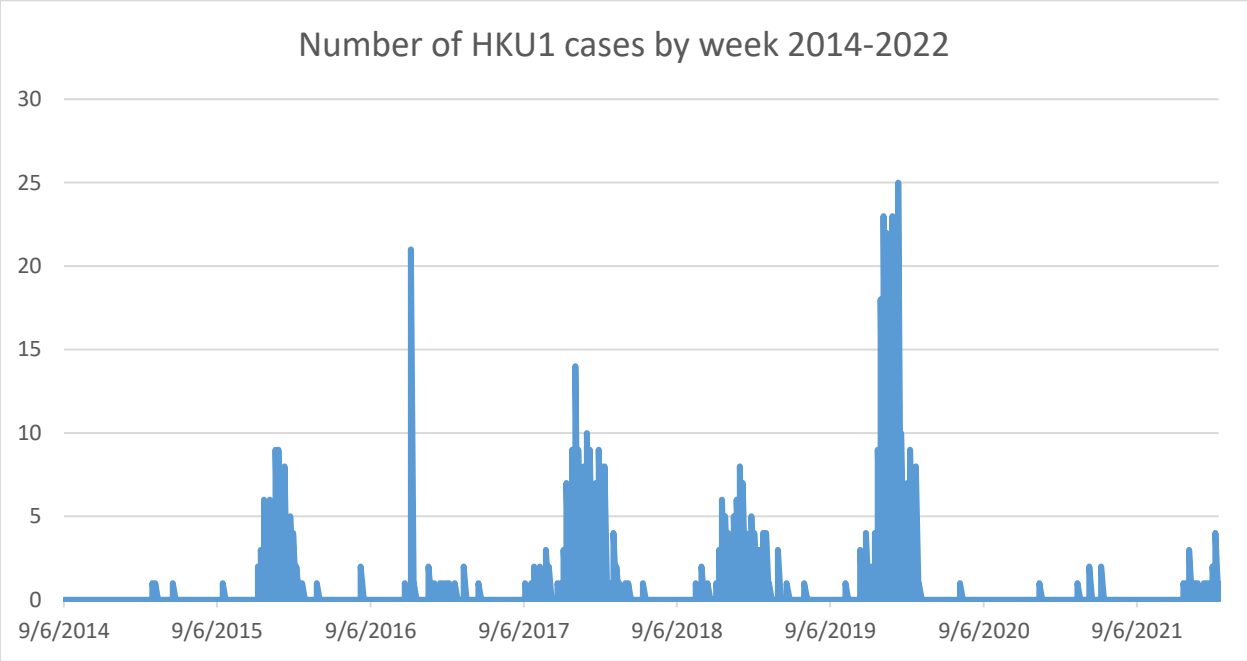
Figure 6. Percentage of children 1-16 years old in Southwestern Pennsylvania positive for 1, 2, 3, or 4 HCoVs, April-July 2020 (n=112) 15

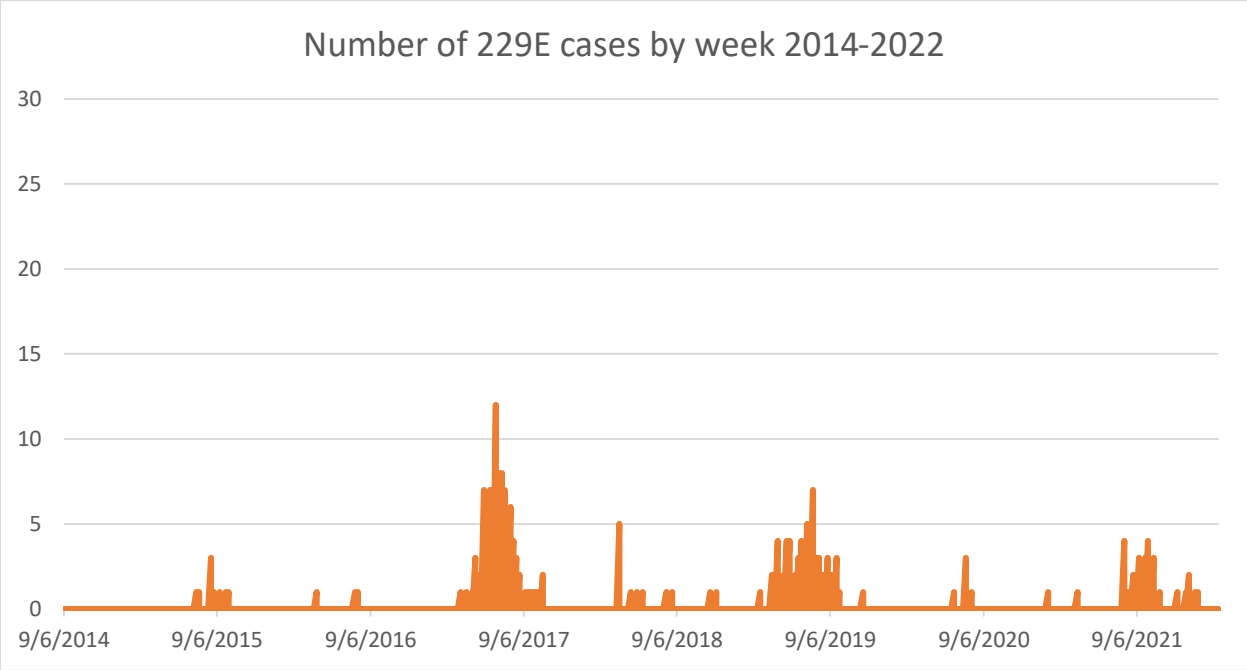
Figure 7. Optical densities for each virus by age group from children 1-15 years old in Southwestern Pennsylvania, April-July 2020 (n=112) 16

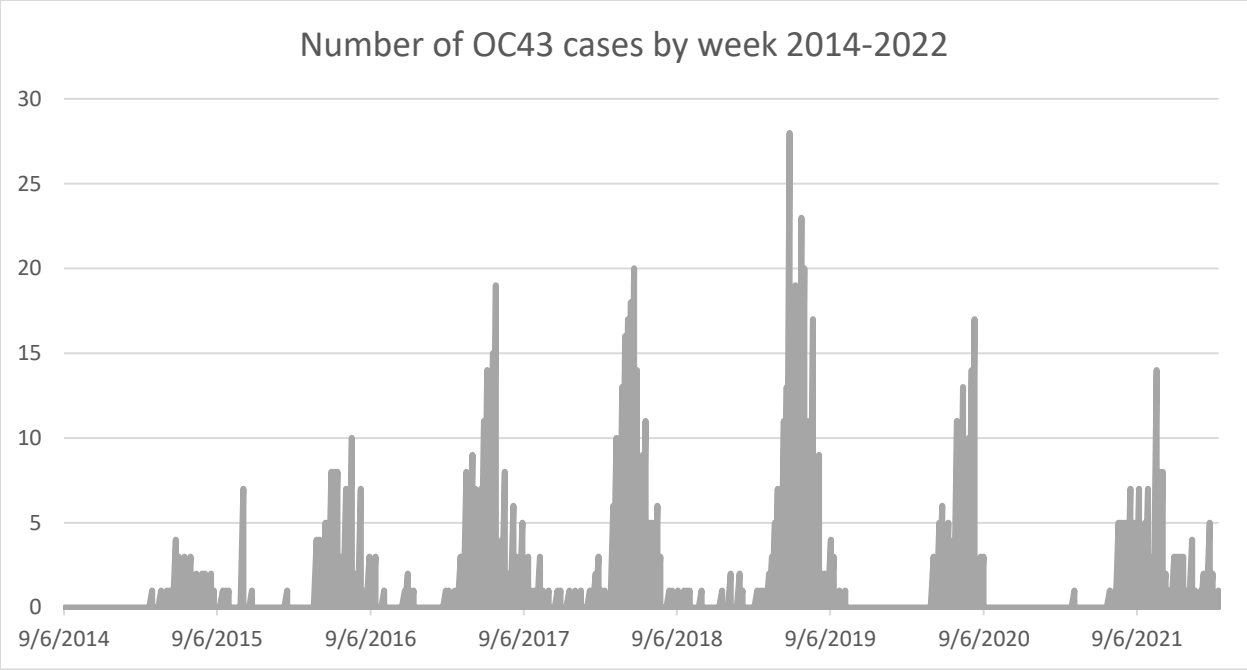
1.0 Background

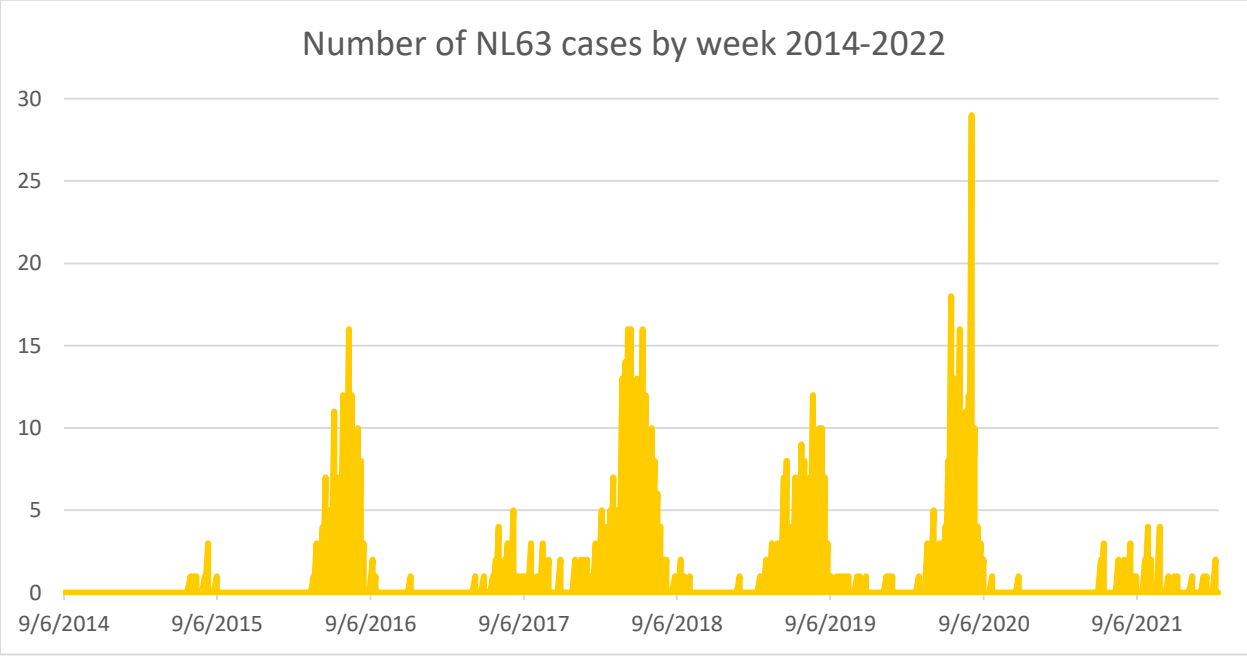
OC43, HKU1, NL63, and 229E are coronaviruses endemic to the United States. These viruses can infect the upper respiratory tract of humans, causing mostly mild but sometimes severe disease. These viruses are single-stranded positive-sense RNA viruses. OC43, HKU1, NL63, and 229E are enveloped viruses and have a capsid surrounding its RNA which protects the genome from damage. The nucleocapsid comprises of RNA and the surrounding capsid. These coronaviruses have lipid envelopes with spike proteins on the surface; these proteins can attach to a variety of surfaces which allow the virus attach to cell surface receptors and enter cells. Between September 6, 2014 and March 19, 2022, there were 112,389 HCoV tests from Pennsylvania reported to the National Respiratory Enteric Virus Surveillance System (*NREVSS*). 581 (0.52%) were positive for HCoV-HKU1, 782 (0.70%) for HCoV-229E, 1112 (0.99%) for HCoV-OC43 and 237 (0.21%) for HCoV-NL63 (NRVESS, 2022). The number of positive cases in Pennsylvania varied by year as seen in Figure 1. Peaks in infection in HKU1, OC43, and NL63 typically occurred in July between 2014 and 2020 whereas in 2021 the months with the highest numbers of cases were September through November for 229E, OC43, and NL63. 229E had the lowest number of infections compared with the other HCoVs; the highest number of cases in a single week reported between 2014 and 2022 was 12 whereas the highest number of cases was 25 for HKU1, 28 for OC43, and 29 for NL63 (NREVSS, 2022). Data specifically for Southwestern Pennsylvania was not available.

Figure 1. Weekly cases of HKU1, 229E, OC43, and NL63 coronaviruses in Pennsylvania, 2014-2022






Number of OC43 cases by week 2014-2022
30
25
20
15
10
5
0
9/6/2014
9/6/2015
9/6/2016
9/6/2017
9/6/2018
9/6/2019
9/6/2020
9/6/2021


Number of NL63 cases by week 2014-2022
30
25
20
15
10
5
0
9/6/2014
9/6/2015
9/6/2016
9/6/2017
9/6/2018
9/6/2019
9/6/2020
9/6/2021

There is a lack of previous literature on the rates on infection of HCoV-HKU1, HCoV-229E, HCoV-NL63, and HCoV-OC43 in different races and sexes, particularly in children. Severe symptoms like pneumonia are an issue of concern and more common in children and the elderly (Killerby, 2018). While there are some studies showing a difference in prevalence in HCoV infections between African Americans and Whites, more research is needed to confirm these results. This project will expand on the limited literature examining the associations between race and sex on HCoV infection. The results may be useful in finding determinants of disease to inform health practitioners which groups are more at risk of developing infections to each HCoVs.

There are several studies showing younger children have higher rates of infection for all four coronaviruses (Wen, 2022). Infections from all four non-severe acute respiratory syndrome human coronaviruses were shown to take place during childhood. One study in Beijing, China measured anti-S IgG antibodies for these four human coronaviruses in 218 children and 576 adults by using immunofluorescence assay. Among 1 to 3 year old children, the rate of prior infection as determined by seropositivity was 70.15% for HCoV-229E, 62.69% for HCoV-OC43, 58.21% for HCoV-HKU1 and 60% for HCoV-NL63. Anti-S IgM was only found in those less than 14 years of age thus demonstrating the majority of seroconversions to each of the four coronaviruses first occurred in children (Zhou, 2013). Another study looked at HCoV-NL63, OC43, 229E, HKU1 in 1062 children who were infected with HCoVs and hospitalized at the Children's Hospital of Hebei Province from January 2015 to December 2020. HCoV-NL63, OC43, 229E, HKU1 infections were most common in children under three years old with 65.72% of infections occurring in the lower respiratory tract (Wen, 2022). A separate study in Hong Kong, China evaluated 4181 patients who were hospitalized with severe acute respiratory infections and found 87 cases of HCoV. Upper respiratory tract infections due to HKU1 affected mainly young children, with or

without underlying diseases while HKU1 pneumonia mainly affected old people with major underlying diseases. A total of 629 children aged 6 months to 5 years were hospitalized for acute respiratory virus infections during the study period (Lau, 2006). Previous research demonstrates that children are more susceptible to infection and severe disease which highlights the importance of studying the prevalence of these coronaviruses to better inform public health officials of disease burden in Southwestern Pennsylvania.

Previous studies have shown an association between seropositivity and demographic categories. Higher rates of seropositivity were observed in African Americans in a previous coronavirus immunoassay study; 69.6% African Americans tested positive for 229E while 52.7% of Caucasians tested positive in this study (Severance, 2008). Differences in association between infection for each common human coronavirus and sex have also been detected. A study in Scotland found more severe cases of common human coronavirus were detected in males than females. From the years 2005 to 2017, 59.2% of CoV-229E, 55.6% of CoV-OC43, and 59.8% of CoV-NL63 cases detected in primary care were in female patients, whereas 54.7% of CoV-229E, 51.1% of CoV-OC43, and 56.7% of CoV-NL63 cases detected in secondary or tertiary care were in male patients although this may reflect differences in healthcare seeking behavior (Nickbakhsh, 2020). Data provided to the NREVSS showed 50.6% of positive HCoV tests were from males (Killerby, 2018). This background information can help us better understand transmission dynamics of coronaviruses.

The study of the prevalence of common human coronaviruses is also important due to literature suggesting past infections of HKU1, 229E, NL63 and OC43 can provide immunity to SARS-CoV-2, although other studies contradict this. In a cohort of 350 individuals uninfected by SARS-CoV-2, a small proportion had IgG antibodies that could cross-react with the S2 subunit of

the SARS-CoV-2 spike protein. The anti-S2 antibodies from patients with HCoVs had neutralizing activity against both SARS-CoV-2 and SARS-CoV-2 S pseudotypes with a much higher percentage of SARS-CoV-2–uninfected children and adolescents positive for these antibodies compared with adults (Ng, 2020). This pattern may be due to children and adolescents having higher HCoV infection and might explain why children are less likely to have severe disease. Conversely, a separate study found 20 percent of 431 serum samples collected before the COVID-19 pandemic had antibodies that can bind to common human coronaviruses and SARS-CoV-2. These cross-reactive antibodies could not neutralize SARS-CoV-2. The levels of cross-reactive anti-CoV antibodies in the blood samples of patients who were later infected with COVID-19 did not correlate with measures of COVID-19 severity such as the need for hospitalization or ICU care (Anderson, 2021).

This project seeks to answer several questions regarding NL63, HKU1, OC43, and 229E human coronaviruses. What is the prevalence of prior infection for NL63, HKU1, OC43, and 229E coronaviruses in children aged 1-16 in Southwestern Pennsylvania? Is there any association between age, race, or sex on infection for each virus? Is there any association between age, race, or sex on optical density (OD) level detected in patient samples for each virus? Does coronavirus seroprevalence increase with age? One null hypothesis is there is no association between age and seropositivity or optical density for HKU1, 229E, OC43, or NL63. The alternative hypothesis is there is an association between age and seropositivity or optical density for HKU1, 229E, OC43, or NL63. On the variables of race and sex, the null hypothesis is there are no associations between either race or sex on seropositivity or optical density for HKU1, 229E, OC43, or NL63. The alternative hypothesis is there are associations between race and sex on seropositivity or optical density for HKU1, 229E, OC43, or NL63.

2.0 Methods

Samples were taken from patients 16 years old and younger with residential addresses in Southwestern Pennsylvania (SWPA) from a previous study examining the prevalence of SARS-CoV-2. The same samples were repurposed for this project examining the prevalence of NL63, HKU1, OC43, and 229E. This was a convenience sample collected from outpatients who visited the University of Pittsburgh Medical Center Children's Hospital of Pittsburgh (UPMC CHP) for routine care. The first phase of collection was between April 27 and May 19, 2020 and the second phase from June 22 to July 3, 2020. The inclusion criteria were that patients had to be between the ages of 6 months old and 19 years and have a residential ZIP code in any of 11 SWPA counties. SWPA as defined by the Pennsylvania Department of Health includes Allegheny, Armstrong, Beaver, Butler, Cambria, Fayette, Greene, Indiana, Somerset, Washington, and Westmoreland counties. The exclusion criteria were: hospital length of stay > 30 days, and receipt of treatments that alter antibody production or antibody profile (immunoglobulin, rituximab, or bortezomib) in the six months prior to collection. Sociodemographic data was collected and is shown in Table 1. A total of 124 samples were available. Samples from patients between 6 and 9 years old and 11 and 14 years old as well as patients older than 16 were not available. Patients under one year of age were excluded from the analysis due to false positivity from maternal antibodies. 27.42% of samples collected were from inpatients, 55.65% from outpatients, and 16.94% from emergency department patients which indicates a wide range of healthcare seeking behavior is represented in the sample. The prevalence of prior infection was determined with the remaining 112 samples.

Table 1. Sociodemographic data of 112 patients less than 16 years old in Southwestern Pennsylvania, April-July 2020

Sociodemographics	n (%)
Race	
Black	21 (18.75%)
White	80 (70.42%)
Asian	2 (1.79%)
Hispanic	2 (1.79%)
Unknown	7 (6.25%)
Sex	
Female	57 (50.89%)
Male	55 (49.11%)
Age	
1	12 (10.71%)
2	21 (18.75%)
3	20 (17.86%)
4	1 (0.89%)
5	20 (17.86%)
10	21 (18.75%)
15	17 (15.18%)
Type of Patient Care	
Inpatient	32 (28.57%)
Outpatient	60 (53.57%)
Emergency	20 (17.86%)

In-house ELISA assays (see appendix for standard operating procedures) were used to determine positive cases for each patient sample. Assays were optimized to detect IgG against the spike (S) and nucleocapsid (N) proteins of NL63, HKU1, OC43, 229E, respectively. Samples were diluted 1:100; optical density readings were taken at 450nm. To determine whether the assays used had a low degree of cross-reactivity, Pearson's R correlations were performed to determine the correlation between optical densities from each ELISA assay detecting for IgG against spike and nucleocapsid proteins. The same analysis was performed for samples when detecting for IgG against N proteins. Only data from spike protein assays were used to determine results due to a high degree of correlation within N protein assays indicating low specificity. A scatter plot was

created to display age and optical density for each virus. Optical density is positively correlated with the amount of antibodies present, higher OD levels indicate a higher amount of antibody presence in the patients which indicates seropositivity. The optical density cutoff for positive and negative cases were determined by observing where the divergence between high optical density and low optical density lies. The positivity cutoff for HKU1, OC43, and NL63 was based on the mean of the lowest 15% of samples due to the fact that the cluster of samples with low optical densities lie below the 15% cutoff. The positivity cutoff for 229E is based on the mean of the lowest 30% of samples due to the fact that the cluster of samples with low optical densities lie below the 30% cutoff. The positivity cutoffs for each virus was calculated by taking the mean OD of the lowest 15% or 30% of the sample readings plus three times the standard deviation. This ensures a 99% confidence that either 15% or 30% of the samples lie below each respective cutoff.

2.1 Statistical Analysis

Multiple logistic regressions were performed in Prism 9 to determine any association between race, sex, or age on seropositivity for each individual virus. Since some samples were missing demographic data, the analysis was performed with the 101 samples for which complete data was available. Asian and Hispanic patients were excluded from the regression as there were only 2 patients from each group. The statistical analysis was limited to African American and Caucasian patients. The results are reported in odds ratios. Multiple linear regressions were performed to determine any association between race, sex, or age on optical density level for each coronavirus. Age was treated as a continuous variable for both statistical tests. Any associations with age will not account for children between 6 and 9 years old, 11 and 14 years old, and children

older than 16. Race and sex were treated as dummy variables with African American and female as the reference variables. Race sex and age are independent variables and positivity or OD are dependent variables. The multiple linear regressions indicate the association between race, sex, and age on the level of antibody response to each virus while the multiple logistic regressions will indicate association between variables and with positivity. Results were considered significant if the P-value was equal or less than 0.05.

3.0 Results

The correlations between the optical densities of each assay are shown. The correlation between different N protein assays ranges between 0.49 and 0.72. This is higher than the correlation between different S protein assays which range between 0.19 and 0.52 which suggests the assays detecting for IgG against spike proteins are more specific and a better measure of OD.

Figure 2. Pearson's R correlation comparing assays by optical density readings

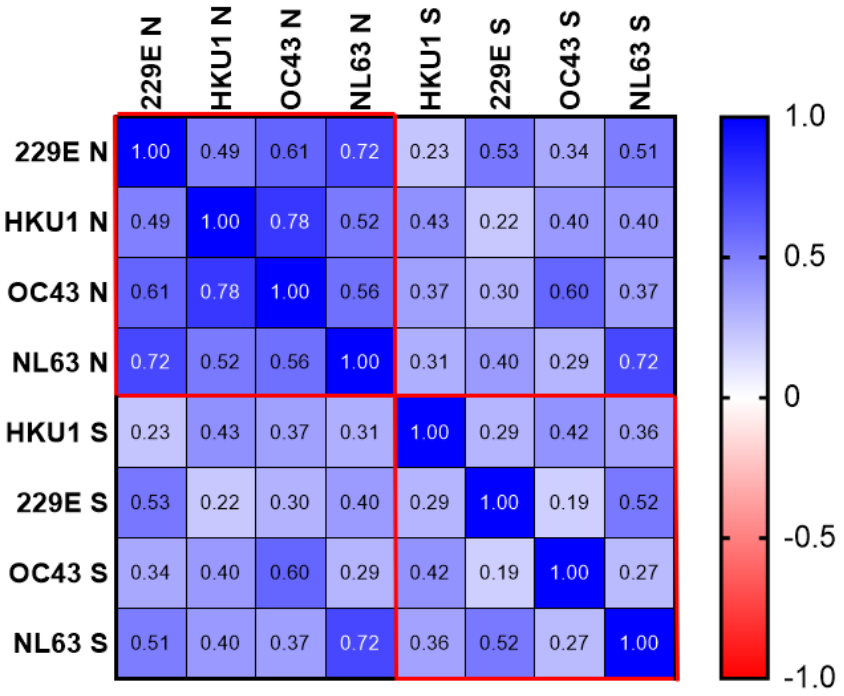


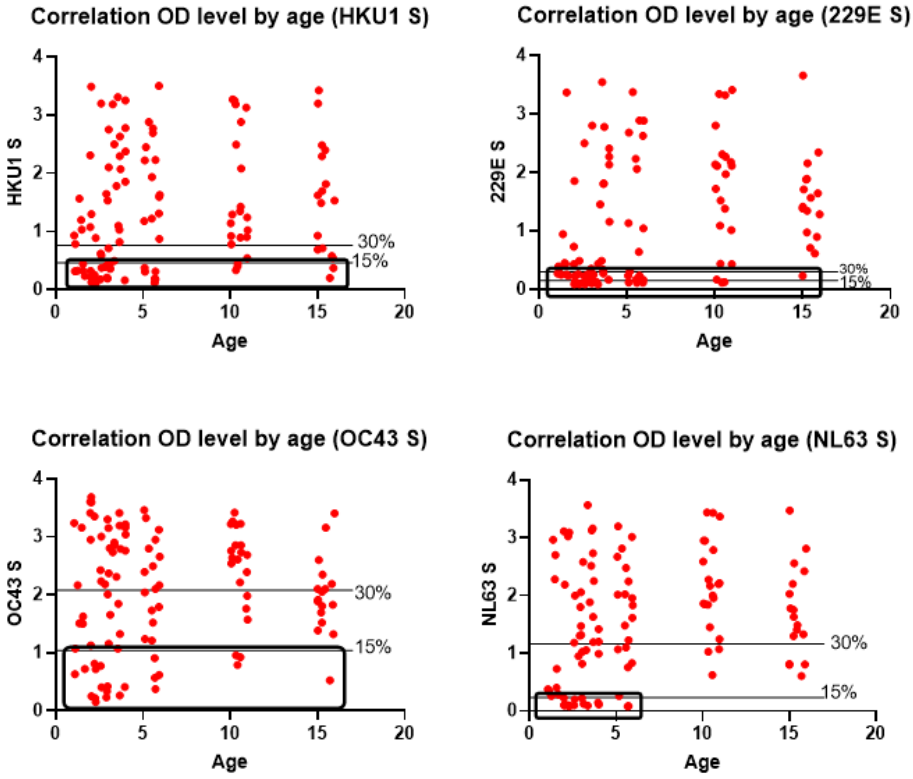
Table 2. R squared values for the correlation of optical densities of ELISAs

	R squared	P value
229E S vs NL63 S	0.27	<0.0001
229E S vs OC43 S	0.04	0.0354
229E S vs HKU1 S	0.08	0.0010
NL63 S vs OC43 S	0.07	0.0026
NL63 S vs HKU1 S	0.13	<0.0001
OC43 S vs HKU1 S	0.18	<0.0001
229E N vs NL63 N	0.52	<0.0001
229E N vs OC43 N	0.37	<0.0001
229E N vs HKU1 N	0.24	<0.0001
NL63 N vs OC43 N	0.31	<0.0001
NL63 N vs HKU1 N	0.27	<0.0001
OC43 N vs HKU1 N	0.61	<0.0001

As shown in Figure 2, the correlations between N protein assays and other N protein assays are higher than the correlations between S protein assays and other S protein assays which suggests a higher specificity in ELISA assays detecting for IgG against S proteins. As shown in Table 2, the R squared values show that a high degree of variation in optical density level of any N protein assay can be explained by the variation in optical density of a second N protein assay. The R squared values are lower when comparing between assays for S protein indicating a lower level of cross reactivity. When using an R^2 cutoff of 0.25 with a P value less than 0.05, only one of the R^2 values for correlations between S protein assays lie above this value. The correlation between the 229E S and NL63 S assay has an R^2 value of 0.27 and a P value less than 0.0001 while all other correlations have R^2 value lower than 0.25. The R^2 value of the correlations between all N protein assays except for the correlation between the 229E N and HKU1 N assays lie above 0.25. The correlation between the 229E N and HKU1 N assays have an R^2 value of 0.24 and a P value less than 0.0001. This indicates a high degree of cross reactivity between antibodies used in the ELISA assays which detected for N proteins. Only data from S protein assays was used in the analysis due to a lower degree of correlation indicating higher specificity.

The scatterplots displaying optical density by age are shown in Figure 3. The positivity cutoff for HKU1, OC43, and NL63 is based on the mean of the lowest 15% of samples. The positivity cutoff for 229E is based on the mean of the lowest 30% of samples.

Figure 3. Representation of patients by age and optical density result



Negative cases are shown inside the box on each graph. Cutoffs were determined by visually delineating where the lowest ODs lie.

3.1 Prevalence of Prior Infection

The prevalence of HCoVs between April 27 and July 3, 2020, was 79.46% for HKU1, 61.61% for 229E, 80.36% for OC43, and 83.93% for NL63 (Figure 4). The prevalence of each

coronavirus in each age group is shown in Figure 5; there was an upward trend in prevalence starting in 2 year old children. 94.11% of 15 year old children were positive for HKU1, 229E, and OC43 indicating that most children will have been infected by their adolescence. All children who were 10 and 15 years of age were positive for NL63. As shown in Figure 6, all children by the age of 1 are positive for at least 2 HCoVs and by age 15, 94.11% are positive for all four HCoVs. The optical density of each sample is shown in Figure 12. 1 and 2 year old children had a lower optical density for HKU1 and 229E than they did for OC43 and NL63 which may indicate a lower circulation of those viruses between 2017 and 2019.

Figure 4. Seroprevalence of HKU1, 229E, OC43, and NL63 coronaviruses in children 1-15 years old in Southwestern Pennsylvania, April-July 2020 (n=112)

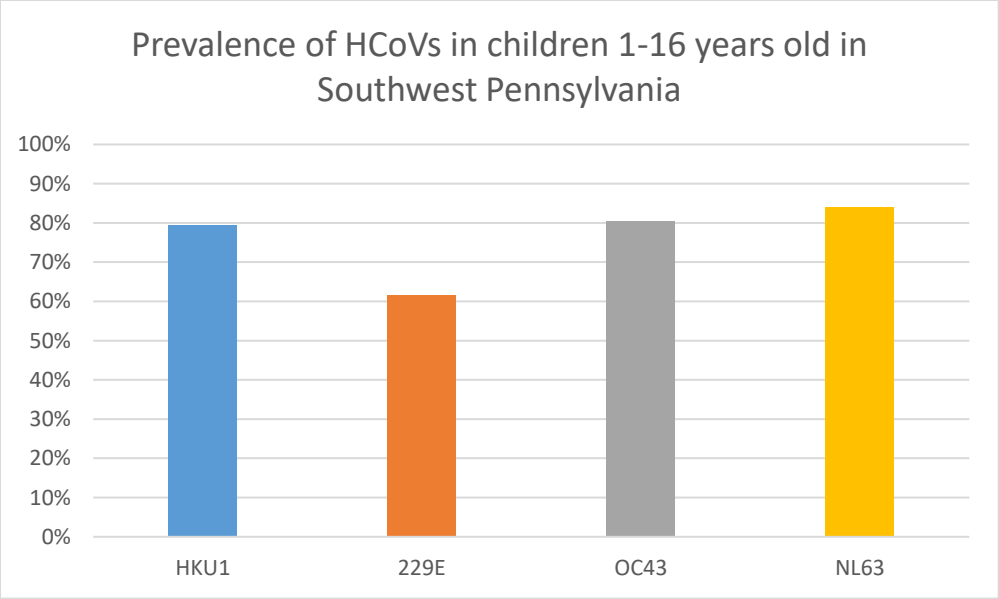


Figure 5. Seroprevalence by age of HKU1, 229E, OC43, and NL63 coronaviruses in children 1-15 years old in Southwestern Pennsylvania, April-July 2020 (n=112)

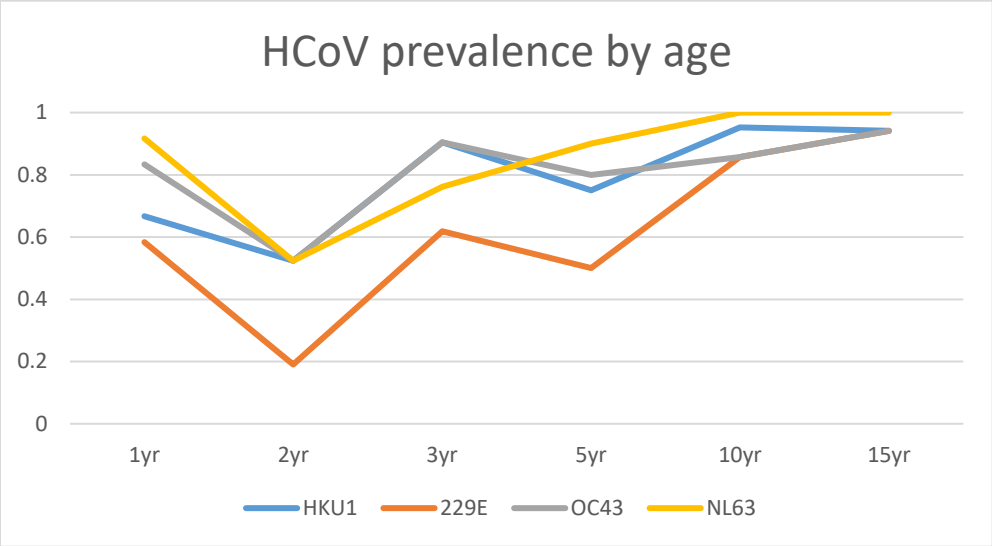
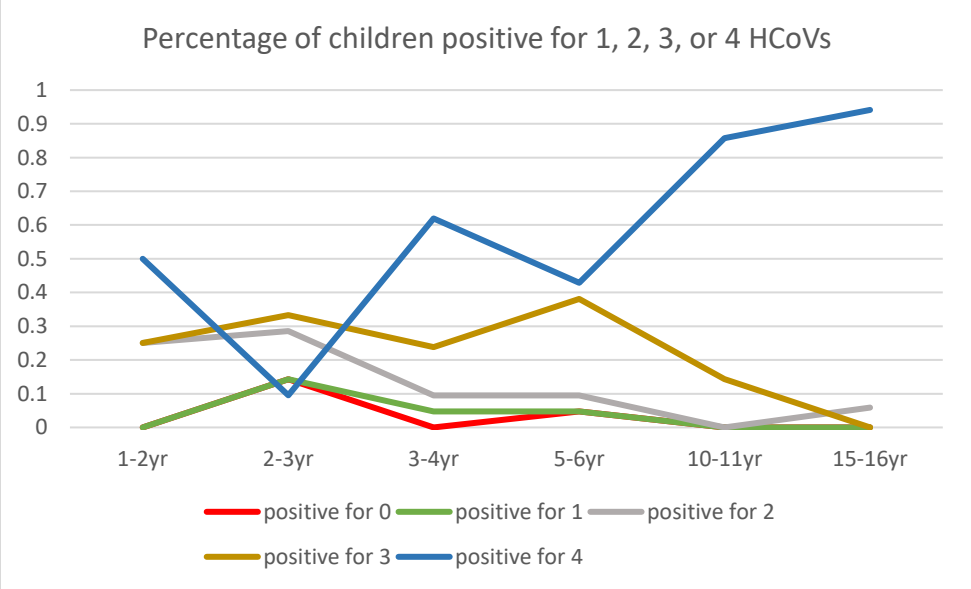
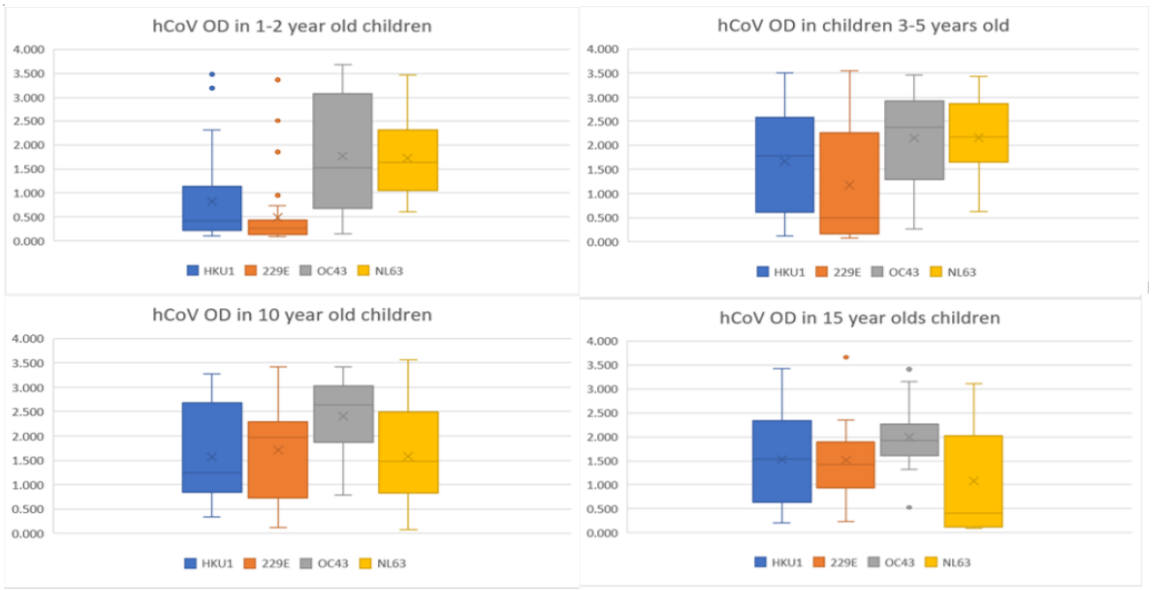


Figure 6. Percentage of children 1-16 years old in Southwestern Pennsylvania positive for 1, 2, 3, or 4 HCoVs, April-July 2020 (n=112)



The percentage of children positive for all 4 coronaviruses increases with age. The percentage of children positive for less than 4 HCoVs is less in older children and 94.11% of 15 year old children are have prior infections of all for HCoVs

Figure 7. Optical densities for each virus by age group from children 1-15 years old in Southwestern Pennsylvania, April-July 2020 (n=112)



3.2 Effect of Variables on Seropositive

Associations were found between age and seropositivity for some viruses as shown in tables 3-6. There was an association between age and seropositivity for the HKU1 coronavirus (Table 3) with an adjusted odds ratio of 0.82 (95 % CI 0.69-0.94; P value = 0.0119) when adjusting for race and sex. This means every 1 year increase in a patient age was associated with a 17.88% lower likelyhood for antibody-positive results for HKU1. There was a significant association between age and seropositivity for 229E shown in table 4. The odds ratio for age was 1.30 (95 %

CI 1.15-1.51; P value = 0.0001) when adjusting for race and sex. This means each one year increase in a patients age was associated with a 30.1% higher likelihood of antibody-positive results for 229E. There was an association between age and seropositivity for NL63. The odds ratio for age is 0.65 (95 % CI 0.45-0.83; P value = 0.0064) (Table 5) when adjusting for race and sex. This means each one year increase in a patients age was associated with a 35% lower likelihood of antibody-positive results for NL63. There was no significant association between age and seropositivity for OC43 (Table 6). The analysis cannot detect for trends in 6-9 and 11-4 year old children because samples from this age were not available. There were no significant associations between sex or race and seropositivity for any of the 4 coronaviruses assessed (Table 3-6).

Table 3. Association between age, race, and sex on HKU1 seropositivity

Odds ratios	**Variable**	**Estimate**	**95% CI (profile likelihood)**
β0	Intercept	0.5935	0.1377 to 2.221
β1	race[Caucasian]	1.129	0.3321 to 4.564
β2	sex[Male]	1.273	0.4514 to 3.662
β3	age	0.8212	0.6867 to 0.9404
Sig. diff. than zero?	**Variable**	**|Z|**	**P value**
β0	Intercept	0.7525	0.4517
β1	race[Caucasian]	0.1855	0.8529
β2	sex[Male]	0.4562	0.6482
β3	age	2.515	0.0119

Table 4. Association between age, race, and sex on 229E seropositivity

Odds ratios	Variable	Estimate	95% CI (profile likelihood)
β0	Intercept	0.2796	0.07534 to 0.9338
β1	race[Caucasian]	1.221	0.3932 to 3.760
β2	sex[Male]	1.209	0.4811 to 3.075
β3	age	1.301	1.153 to 1.511
Sig. diff. than zero?	**Variable**	**|Z|**	**P value**
β0	Intercept	2.014	0.0440
β1	race[Caucasian]	0.3498	0.7265
β2	sex[Male]	0.4044	0.6859
β3	age	3.872	0.0001

Table 5. Association between age, race, and sex on NL63 seropositivity

Odds ratios	Variable	Estimate	95% CI (profile likelihood)
β0	Intercept	1.224	0.2164 to 6.376
β1	race[Caucasian]	1.752	0.4191 to 9.526
β2	sex[Male]	0.4490	0.1259 to 1.480
β3	age	0.6500	0.4460 to 0.8330
Sig. diff. than zero?	**Variable**	**|Z|**	**P value**
β0	Intercept	0.2423	0.8086
β1	race[Caucasian]	0.7209	0.4710
β2	sex[Male]	1.290	0.1969
β3	age	2.727	0.0064

Table 6. Association between age, race, and sex on OC43 seropositivity

Odds ratios	Variable	Estimate	95% CI (profile likelihood)
β0	Intercept	0.5422	0.1311 to 1.923
β1	race[Caucasian]	1.245	0.3766 to 4.977
β2	sex[Male]	0.5759	0.1994 to 1.585
β3	age	0.8901	0.7751 to 0.9981
Sig. diff. than zer	**Variable**	**|Z|**	**P value**
β0	Intercept	0.9147	0.3604
β1	race[Caucasian]	0.3392	0.7344
β2	sex[Male]	1.055	0.2915
β3	age	1.840	0.0658

As shown in Table 7 for every 1 year increase in patient age there was a 0.07 increase in optical density when detecting for IgG against 229E S proteins (P value = 0.0007, 95% CI 0.03 to 0.12). For every 1 year increase in patient age there was a 0.06 increase in optical density when detecting for IgG against NL63 S proteins (P value = 0.0091, 95% CI 0.01 to 0.10) as seen in table 8. This is in contrast with the results from the multiple logistic regression which showed a decrease in likelihood that a patient was seropositive for NL63 as they aged (Table 6). As seen in Table 9-10, there is no association between age and optical density measuring IgG against OC43 or HKU1. The analysis cannot detect for trends in 6-9 and 11-4 year old children because samples from this age were not available. There was no significant association between sex and race and OD level for any of the four viruses.

Table 7. Results of the multiple linear regression with race, sex, and age as independent variables and optical density measuring IgG against 229E as the dependent variable

Variable	Estimate	95% CI (asymptotic)	P value
Intercept	0.56	0.01 to 1.10	0.0458
race[Caucasian]	0.11	-0.40 to 0.61	0.6678
sex[Male]	0.00	-0.41 to 0.40	0.9819
age	0.075	0.03 to 0.12	0.0007

Table 8. Results of the multiple linear regression with race, sex, and age as independent variables and optical density measuring IgG against NL63 as the dependent variable

Variable	Estimate	95% CI (asymptotic)	P value
Intercept	1.10	0.56 to 1.63	<0.0001
race[Caucasian]	-0.07	-0.56 to 0.43	0.7910
sex[Male]	0.34	-0.06 to 0.74	0.0922
age	0.06	0.01 to 0.10	0.0091

Table 9. Results of the multiple linear regression with race, sex, and age as independent variables and optical density measuring IgG against HKU1 as the dependent variable

Variable	Estimate	95% CI (asymptotic)	P value
Intercept	1.27	0.75 to 1.80	<0.0001
race[Caucasian]	-0.20	-0.69 to 0.28	0.4047
sex[Male]	-0.10	-0.49 to 0.30	0.6309
age	0.04	0.00 to 0.08	0.0647

Table 10. Results of the multiple linear regression with race, sex, and age as independent variables and optical density measuring IgG against OC43 as the dependent variable

Variable	Estimate	95% CI (asymptotic)	P value
Intercept	1.84	1.30 to 2.37	<0.0001
race[Caucasian]	-0.02	-0.52 to 0.47	0.9204
sex[Male]	0.25	-0.15 to 0.65	0.2211
age	0.02	-0.02 to 0.06	0.3951

4.0 Discussion

This project sought to answer two questions. What is the prevalence of prior infection for NL63, HKU1, OC43, and 229E coronaviruses in children aged 1-16 in Southwestern Pennsylvania? Is there any association between age, race, or sex on infection for each virus?

Based on the results of my analysis, I reject the null hypotheses that there is no association between age and seropositivity or optical density for HKU1, 229E, OC43, and NL63 For every 1 year age increase, the likelihood of seropositivity for HKU1 decreases by 17.88% when adjusting for race and sex. Each one year increase in a patients age was associated with a 30.1% higher likelihood of antibody-positive results for 229E when adjusting for race and sex and for every 1 year increase in patient age there is a 0.08 increase in optical density when detecting for IgG against 229E S proteins. Each one year increase in a patients age was associated with a 35% lower likelihood of antibody-positive results for NL63 when adjusting for race and sex and for every year increase in patient age there is a 0.06 increase in optical density when detecting for IgG against NL63 S proteins. I faill to reject the second null hypothesis that there is no association between race or sex on seropositivity or optical density for HKU1, 229E, OC43, or NL63.

There was a discrepancy in the results for NL63; an odds ratio of 0.65 between age and seropositivity indicates older children hada lower likelihood of being positive. A parameter estimate of 0.06 between age and optical density demonstrates an increase in OD with age, indicting a greater antibody response in older children. This discrepancy may be due to three factors: an incorrect cutoff between positive and negative cases thus leading to more cases being considered negative in older children, high cross reactivity between the ELISA assay for the NL63 spike and other assays, or sampling bias since no samples from children 6 to 9 and 11 to 14 years

old were available. It was also seen that the prevalence of prior infection is higher in 10 and 15 year old children which appears to contradicts the odds ratios between age and seropositivity for HKU1 and NL63. This may be due to the lack of samples from 6-9 and 11- 14 year old children.

The assumption that older children are more likely to be positive for one of the four coronaviruses is supported by the results, which showed all children by the age of 1 are positive for at least 2 HCoVs and by age 15, 94.12% are positive for all four HCoVs. Results from logistic regressions predicting the effects of age on seropositivity for 229E and results from the linear regressions predicting effects of age on OD for 229E and NL63 support this conclusion. As age increases, optical density for 229E and NL63 do as well. However, results from logistic regressions predicting effects of age on seropositivity for NL63 and HKU1 show that older children had a lower likelihood of being positive for each of these viruses. Age may be a determinant of disease for infections of HKU1, 229E, and NL63. Age does not appear to affect the likelihood of infection for OC43. Race and sex were, however, not associated with likelihood of infection for any of the four coronaviruses. The optical density levels of 1-2 year old children for HKU1 are lower than for OC43 and NL63, which is consistent with the state aggregate data for weekly cases in Pennsylvania between 2017 and 2019. This indicates a low level of circulation of HKU1 specific to Southwestern Pennsylvania.

The findings of this project can inform health practitioners and public health officials on the circulation of common human coronaviruses in Southwestern Pennsylvania. There were no previous studies looking at HKU1, 229E, OC43, and NL63 infection in children in this region. Common human coronaviruses cause more severe symptoms in children and the elderly and thus it is useful to know trends in infection. Unlike previous literature, this project did not find significant associations for sex or race on OD level or positivity, although it is possible the sample

is too small to detect an association. The finding that 94.11% of 15-year old children were positive for HKU1, 229E, and OC43, and that 100% of all 10-15 year old children were positive for NL63, suggests that coronavirus surveillance should focus on children.

There are several limitations to the analysis, the most important of which is the sample size. The number of samples used for the analysis is 101, which is likely underpowered to assess significance for small effect sizes. This study is also limited by the lack of confirmed negative samples. Due to the fact that each four HCoV are common viruses, most children will have already been exposed to the viruses and thus there are no definitive negative controls. As a result the cutoffs for positive and negative cases were determined by observing where the divergence between high optical density and low optical density which may be prone to error. Patient behavior such as hygeine and healthcare seeking behavior was also not considered and may be a factor on which groups are more likely to become infected with HCoVs. Hygeine behavior such as hand washing and wearing masks were not tracked and may have influenced the results of the regression analyses since people who wear masks and wash their hands frequently are less likely to be infected. Another limitation is the lack of information on antibody persistence for each of the four coronaviruses. Previous literature has not shown a consistent time frame for antibody persistence in children; therefore the time of the most recent infection can affect the analysis involving antibody level. Furthermore there were no samples taken from children between the ages of 6-9 and 11-14 and all samples where taken from children undergoing routine clinical care to Children's Hospital of Pittsburgh. This sample does not account for children who did not visit the hospital and may not be representative of the population of Southwestern Pennsylvania. A more representative sample set is needed for future studies. Future studies should attempt to recruit

patients from both inside and outside of the hospital setting, recruit a larger number of patients from all ages, and track hygene behavior.

Appendix SOP: Total IgG spike and nucleocapsid of common cold coronaviruses-indirect ELISA

Elaborated by:	Reviewed by:	Approved by:
Name: Priscila Castanha Position: Postdoctoral Associate	Name: Position:	Name: Position:
		Effective since: June 2021

Glossary and Abbreviations

CCC – Common cold coronaviruses

S – Spike protein

NP – nucleoprotein

1. Required Reagents and Equipment

2. Reagents:

6.1.1 Recombinant proteins

Appendix Table 1. Recombinant Proteins

Virus	Target	Manufacturer	Expression Host	cat#
HCoV-HKU1	HKU1 (isolate N1) Spike/S1 Protein (S1 Subunit, His Tag)	Sino Biological	HEK293 Cells	40021-V08H
HCoV-NL63	NL63 Spike/S1 Protein (S1 Subunit, His Tag)	Sino Biological	HEK293 Cells	40600-V08H

HCoV-229E	229E Spike/S1 Protein (S1 Subunit, His Tag)	Sino Biological	HEK293 Cells	40601-V08H
HCoV-OC43	OC43 Spike Protein (S1+S2 ECD, His Tag)	Sino Biological	Baculovirus-Insect Cells	40607-V08B
SARS-COV	SARS CoV Spike Protein (delta-TM*)	BEI	Baculovirus-Insect Cells	NR-722
HCoV-HKU1	HKU1 Nucleocapsid Protein (His Tag)	Sino Biological	E. coli	40642-V07E
HCoV-NL63	NL63 Nucleocapsid Protein (His Tag)	Sino Biological	E. coli	40641-V07E
HCoV-229E	229E Nucleocapsid Protein (His Tag)	Sino Biological	E. coli	40640-V07E
HCoV-OC43	OC43 Nucleocapsid Protein (His Tag)	Sino Biological	E. coli	40643-V07E

6.1.2 Goat Fab_2 anti human IgG (H+L)-HRP (Jackson Immunoresearch, cat# 109-036-003; lot: 106902)

6.1.3 Blotting-Grade Blocker, 300g (Biorad; cat# 170-6404)

6.1.4 Boston Bioproducts Inc PBS 10X PH7.4 1000mL (Fisher; cat# NC9140736)

6.1.5 SureBlue Reserve™ TMB Microwell Peroxidase Substrate (KPL; cat# 53-00-03)

6.1.6 HCl (Sigma; cat# H1758-500ML)

6.1.7 Tween 20 (Fisher, cat# BP337-500)

6.1.8 Costar 96-Well Half-Area Plates, ELISA Plate/ flat bottom High binding 100/cs (Fisher; cat# 07-200-37)

3. Equipment

6.2.1 Molecular Devices SpectraMax Plus 384 microplate reader

6.2.2 Biotek Plate washer 405S model

4. Solutions

6.3.1 Washing buffer – 1X PBS; 0.1% (v/v) Tween 20

6.3.1.1 1000mL of PBS 10X + 10mL of Tween 20 + 8990mL of deionized water)

6.3.2 Blocking buffer and Dilution buffer – 5% and 1% (w/v) Blotting-Grade Blocker in washing buffer, respectively

6.3.2.1 Blocking: 5g of Blotting-Grade Blocker + 100mL of washing buffer

6.3.2.2 Dilution: 1g of Blotting-Grade Blocker + 100mL of washing buffer

6.3.3 Carbonate/bicarbonate buffer (BupH Carbonate-Bicarbonate Packs; Pierce, cat# 28382)

6.3.3.1 500mL of deionized water per pack

6.3.4 Goat Fab_2 anti human IgG (H+L)-HRP at 1:20,000

6.3.4.1 1uL of anti-human IgG + 19,999uL of dilution buffer per ELISA plate

5. Procedure

7.1 Coat half area ELISA plate with 50uL per well of the recombinant protein of interest at 2ug/mL dilution in carbonate/bicarbonate buffer and incubate for overnight (16hs-18hs) at 4^{o}C in a humid chamber

7.2 Aspirate plate ["ASP" program (no. 19) on the plate washer] and add 150uL per well of blocking buffer per well. Incubate for 1 hour at room temperature (22^{o}C$\pm$2)

7.3 Aspirate plate ["ASP" program (no. 19) on the plate washer)] and add 50uL per well of serum or plasma samples and assay controls diluted at 1:100 in dilution buffer. Incubate plate for 1hr at room temperature (22^{o}C$\pm$2)

7.4 Wash 6 times with washing buffer using "STANDARD 6X" program (no. 19) on the plate washer and add 50uL per well of Goat Fab_2 anti human IgG (H+L)-HRP diluted at 1:20,000 in dilution buffer. Incubate for 1hr at room temperature (22^{o}C$\pm$2)

7.5 Wash 4 times with washing buffer using "F1NAL wash" program on the plate washer (no. 20). In the end the plate wash will add 190uL of washing buffer. Let plate soaking for 5

minutes and then aspirate plate ["ASP" program (no. 19) on the plate washer)] and add 50uL per well of TMB. Incubate for 20 minutes at room temperature ($22^{o}C \pm 2$).

7.6 Stop reaction with 50uL per well of 1N HCl

7.7 Read plate at 450nm

7.8 Analysis

If a sample's average OD450nm is above the cut-off value for the plate, it will only be considered eligible for analysis if coefficient of variation for the replicates is below 20%

7.8.1 Subtract OD samples from blank from this point on

7.8.2 Calculate cut-off based on the following equation

Cut-off = Avg Negative Control + 3x(STDEV Negative Control)

7.8.3 If a sample's average OD450nm is above the cut-off value for the plate, it will only be considered eligible for analysis if coefficient of variation for the replicates is below 20%

8 Important notes

8.1 Add first plate controls and then samples

8.2 Prepare Goat Fab_2 anti human IgG (H+L)-HRP dilution within the last 20 minutes of the serum incubation

8.3 Let an aliquot of TMB to warm up for 1 hour at room temperature before adding to the ELISA plate

1. Safety information

Using coats and gloves is mandatory

Bibliography

1. Anderson, E. M., Goodwin, E. C., Verma, A., Arevalo, C. P., Bolton, M. J., Weirick, M. E., Gouma, S., McAllister, C. M., Christensen, S. R., Weaver, J., Hicks, P., Manzoni, T. B., Oniyide, O., Ramage, H., Mathew, D., Baxter, A. E., Oldridge, D. A., Greenplate, A. R., Wu, J. E., Alanio, C., … Hensley, S. E. (2021). Seasonal human coronavirus antibodies are boosted upon SARS-CoV-2 infection but not associated with protection. *Cell*, *184*(7), 1858–1864.e10.

2. Killerby, M. E., Biggs, H. M., Haynes, A., Dahl, R. M., Mustaquim, D., Gerber, S. I., & Watson, J. T. (2018). Human coronavirus circulation in the United States 2014-2017. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, *101*, 52–56.

3. Lau, S. K., Woo, P. C., Yip, C. C., Tse, H., Tsoi, H. W., Cheng, V. C., Lee, P., Tang, B. S., Cheung, C. H., Lee, R. A., So, L. Y., Lau, Y. L., Chan, K. H., & Yuen, K. Y. (2006). Coronavirus HKU1 and other coronavirus infections in Hong Kong. *Journal of clinical microbiology*, *44*(6), 2063–2071.

4. National Respiratory and Enteric Virus Surveillance System. (2022). Final_PA Lawrence_Ji. [Data File]. Atlanta: Center for Disease Control and Prevention.

5. Ng, K. W., Faulkner, N., Cornish, G. H., Rosa, A., Harvey, R., Hussain, S., Ulferts, R., Earl, C., Wrobel, A. G., Benton, D. J., Roustan, C., Bolland, W., Thompson, R., Agua-Doce, A., Hobson, P., Heaney, J., Rickman, H., Paraskevopoulou, S., Houlihan, C. F., Thomson, K., … Kassiotis, G. (2020). Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science (New York, N.Y.)*, *370*(6522), 1339–1343.

6. Nickbakhsh, S., Ho, A., Marques, D., McMenamin, J., Gunson, R. N., & Murcia, P. R. (2020). Epidemiology of Seasonal Coronaviruses: Establishing the Context for the Emergence of Coronavirus Disease 2019. *The Journal of infectious diseases*, *222*(1), 17–25.

7. Severance, E. G., Bossis, I., Dickerson, F. B., Stallings, C. R., Origoni, A. E., Sullens, A., Yolken, R. H., & Viscidi, R. P. (2008). Development of a nucleocapsid-based human coronavirus immunoassay and estimates of individuals exposed to coronavirus in a U.S. metropolitan population. *Clinical and vaccine immunology : CVI*, *15*(12), 1805–1810.

8. Wen, C., Sun, L., Zhao, M. C., Duan, S. X., Wang, L., & Cui, X. W. (2022). Clinical Study of Human Coronavirus NL63, OC43, 229E, HKU1 Infentions in Hospitalized Children from 2015 to 2020. *Infection and drug resistance*, *15*, 1093–1101.

9. Zhou, W., Wang, W., Wang, H., Lu, R., & Tan, W. (2013). First infection by all four non-severe acute respiratory syndrome human coronaviruses takes place during childhood. *BMC infectious diseases*, *13*, 433.