General Introduction:
We used the published research methods to estimate the overall number of Latinos in Washington state affected by Long COVID.

We reference the methods used in the Colorado Long COVID Report. The specific methodologies they used are on the website on page 16: https://www.documentcloud.org/documents/23597651-colorado-long-covid-report-01032023-final-accessible-1.

We used the published statistical estimate coefficients to estimate Latinos affected by Long COVID; using these two published research methods but making slight adjustments based on the conditions in Washington state:
The method for adult estimation was originally developed by Chen et al. and published in the Journal of Infectious Disease in 2022. https://pubmed.ncbi.nlm.nih.gov/35429399/

The method for pediatric estimation was originally developed by Lopez-Leon et al and published in Scientific Reports. https://www.nature.com/articles/s41598-022-13495-5

The method for the elders' estimation was a variation from the adult estimation, in which we used the coefficients of adult estimation times 1.331.

We used two datasets obtained from the Washington state COVID-19 Data Dashboard. https://doh.wa.gov/emergencies/covid-19/data-dashboard#CurrentStatus

Processing First Dataset:

The first dataset counts the total positive cases and hospitalized conditions by different age groups. We categorized the age group dataset into smaller subsets, mainly three categories: Under 18, 18 to 49, and 50+.

We summarized the total cases and hospitalized cases, then used total cases minus hospitalized cases to get the total non-hospitalized cases. Because the dataset counts the cases using a 7 days period, we divided the total values by 7 to get the true cases number. After the process, we have the true total cases, hospitalized cases, and non-hospitalized cases for the entire population of Washington state.
We then get the Latinos cases by using the percentage of the Latinos population in Washington state, which is 13.4%. We time 13.4% of the population to get the Latinos cases estimation.

We then use the coefficients, with an estimation range to get the estimation.

For hospitalized pediatrics, we multiply the population by 0.29(low: 0.18 / high: 0.81). For non-hospitalized pediatrics, we multiply the population by 0.25(low: 0.18 / high: 0.33).

For hospitalized adults, we multiply the population by 0.54(low: 0.44 / high: 0.63). For non-hospitalized adults, we multiply the population by 0.34(low: 0.25 / high: 0.46).

For hospitalized elders, we multiply the population by 0.72(low: 0.59 / high: 0.84). For non-hospitalized elders, we multiply the population by 0.45(low: 0.33 / high: 0.61).

Processing Second Dataset:

The second dataset counts the total positive cases and hospitalized conditions by different genders. We categorized the age group dataset into smaller subsets, male and female.

We follow the same process as we did in the first dataset. We summarized the total cases and hospitalized cases then used total cases minus hospitalized cases to get the total non-hospitalized cases. Because the dataset counts the cases using a 7 days period, we divided the total values by 7 to get the true cases number. After the process, we have the true total cases, hospitalized cases, and non-hospitalized cases for the entire population of Washington state.
We then get the Latinos cases by using the percentage of the Latinos population in Washington state, which is 13.4%. We time 13.4% of the population to get the Latinos cases estimation.

We then use the coefficients, with an estimation range to get the estimation.

For hospitalized males, we multiply the population by 0.54(low: 0.44 / high: 0.63). For the non-hospitalized males, we multiply the population by 0.37(low: 0.25 / high: 0.51).

For hospitalized females, we multiply the population by 0.54(low: 0.44 / high: 0.63). For the non-hospitalized females, we multiply the population by 0.43(low: 0.35 / high: 0.63).

Potential Biases

This estimation has several potential biases that might lead to estimation errors.

The first potential bias is that the coefficients we used for estimation come from the published research based on the age group and gender group. They did not take ethnicity as a considerable factor. However, during our research, there is solid evidence that minority and ethnicity could be an important factor that affects the probability of getting Long COVID. Here we propose a hypothesis that the true coefficients for the Latinos community will be higher than the coefficients we used. Therefore, our estimation could just be a baseline and minimal estimation for the affected population of Latinos.

The second potential bias is the coefficient for the elders' estimation, which used the adult coefficient times 1.331. However, 1.331 is the factor Colorado Long COVID Report gave, and it does not include any scientific research or publications to account for this number. This could be the specific value just for Colorado state. So we should do further research and conclude a coefficient for elders that could be used in Washington state elders estimation.

For the elders coefficient testing, we will also use similar statistics to validate if the coefficient will be the adult coefficient times 1.331. We propose our null hypothesis with H0: c=1.331 and do a sample testing choosing participants whose ages are over 50. Then we could present a T-test to see if we could support our null hypothesis or reject it.

Further Validation:

To validate our estimations, there are some steps we could dive deeper into.

We could validate our estimations for Latinos communities by taking samples from the specific counties in the Eastern Washington area. The sample testing involves a group selection that contains a certain number of people and counts the total number of the population who have Long COVID. The selection of the sample should be independent each time so we could get a result with the lowest standard error.

In the statistical description, instead of counting the affected population in total, we can select counties that contain a minimum population of 20% of the total Latinos population to do sample testing. The general process of the sample testing is that we first calculate the mean Latinos affected ratio from our estimation, which is the Long COVID-affected estimated population divided by the whole Latino population. Then we want to test our null hypothesis, which is whether the mean ratio of our estimation is the same as the mean ratio of our sample testing. For the sample testing, we will divide the sample counties population into N groups, each group containing M numbers of participants. So there are total NxM participants in this sample testing. By performing a T-test, we can calculate a value called p-value that can validate our null hypothesis. If the p-value is below 0.05, we can say that our null hypothesis cannot be rejected, which means our estimation has high precision. If the p-value is over 0.05, we have to step back and reconsider our coefficients and think of other factors that might lead to different estimation results.

References:

"Colorado State." *DocumentCloud*, https://www.documentcloud.org/documents/23597651-colorado-long-covid-report-01032023-final-accessible-1.

Chen C;Haupert SR;Zimmermann L;Shi X;Fritsche LG;Mukherjee B; "Global Prevalence of Post-Coronavirus Disease 2019 (COVID-19) Condition or Long Covid: A Meta-Analysis and Systematic Review." *The Journal of Infectious Diseases*, U.S. National Library of Medicine, https://pubmed.ncbi.nlm.nih.gov/35429399/.

Lopez-Leon, Sandra, et al. "Long-Covid in Children and Adolescents: A Systematic Review and Meta-Analyses." *Nature News*, Nature Publishing Group, 23 June 2022, https://www.nature.com/articles/s41598-022-13495-5.

"Covid-19 Data Dashboard." *Washington State Department of Health*, https://doh.wa.gov/emergencies/covid-19/data-dashboard#CurrentStatus.