Assessing COVID-19 Severity in US Children with Bayesian Random Time Effect Model

Sian Chee Tong

May 3, 2022

Abstract

The novel coronavirus disease 2019 (COVID-19) has been widely reported in all age groups including children, adults and elderly populations in US. However, relatively little attention across COVID-19 studies were made on understanding the associativity of factors affecting the disease severity among children.

Using the COVID-19 Case Surveillance Public Use Data [4], we assess the factors associated with clinical severity (hospitalization, intensive care unit (ICU) admission, and deaths) among the children who contracted COVID-19. Descriptive exploratory analyses were conducted using the subset of the data for children and using observations since March 1 to December 16, 2020. Patterns of missing data in the dataset were addressed. Bayesian random time effect Poisson models were implemented to estimate the cumulative weekly adjusted COVID-19-associated incidence rates for all severities. Comparison between demographic (sex and race) and clinical (presence of underlying medical conditions) characteristics among the children patients were made using incidence rate ratios.

Our results revealed that race and ethnicity disparities were noticeably seen among the severity outcomes in children patients. Black and Hispanic/Latino children severity outcomes were estimated to be disproportionately higher than White children which are consistent among many prior studies. Severity outcomes among children with underlying medical conditions were at least three times more likely compared to those without. We also find that gender did not provide significant contribution in explaining the severity outcomes.

1 Introduction

The COVID-19 has caused a widespread in cases and deaths worldwide since January 2020 and United States (US) is one of the worse hit country by the pandemic [9]. Although infections is usually non-fatal, severe COVID-19 can present life-threatening illness ([10] and [20]).

As many studies have widely established understanding towards the severity of COVID-19 infections on adults and elderly age groups, more research and evidence is required for evaluating the factors that affected children patients. This is complicated by difference in results obtained from different sources. According to the American Academy of Pediatrics [16], a total of 2,000,681 child COVID-19 cases has been reported representing 12.4% of all cases in the US by December 14, 2020. By April 2022, children less than 18 years old account for around 18 to 19% of the cumulative laboratory-confirmed cases [6]. Although children typically has lower risk of exposure and tested less frequently compared to adults, the incidence rate of children was similar to adults [5]. Further, available evidence suggested that children who contracted COVID-19 are not spared from developing long-term effects of the disease [19].

Therefore, careful monitoring of COVID-19 infections and severe outcome among childrens remains particularly important. This is to support the ongoing evaluation of public health control and prevention strategies on for instance, school reopening or remote learning and provide more concrete public health guidance towards parents and caregivers in the US.

Thus, the above aforementioned reasons motivated us into studying and re-analyzing the factors that were highlighted by many studies beforehand to gain a comprehensive picture of what affects the severity outcomes among the COVID-19 cases in US children using the COVID-19 case surveillance Public Use Data shared publicly by Centers for Disease Control and Prevention (CDC). Ultimately, we hope that this will support the wider action to protect childrens and citizens from COVID-19.

2 Background

A few recent works were conducted using similar datasets on discovering relationships among factors that affected COVID-19 infection among children patients.

One of the commonly used dataset is the COVID-19 Associated Hospitalization Surveillance Network (COVID-NET) [7]. The COVID-NET provides race and ethnicity information for hospitalization among children Using the dataset, Kim et al. [11] reveals that one in a third of children with COVID-19 were admitted to ICU. The study found that weekly hospitalization rates in children has been increasing during the surveillance period. Among them, Hispanic and Black children had the highest cumulative hospitalization rates compared to White children.

In one of CDC's morbidity and mortality weekly report [1], COVID-19 death data among persons < 21 years old were collected from public health jurisdictions in US between February 12 to July 31 2020. They found that among 121 COVID-19-associated deaths in the US, 63% of the cases were male, 10% of decedents were infants, 20% were aged 1-9 years and 70% were aged 10-20 years. Additionally, it was shown that nearly 75% of the death cases among children less than 21 years old have occurred within age 10-20 years, with a disproportionate percentage among Hispanics, Blacks, American Native/Alaska Natives (AI/AN), and children with underlying medical conditions. This suggested the existence of death disparities among children across gender, age group, and presence of underlying medical conditions factors.

Additionally, a cohort study [14] using the Premier Healthcare Database Special COVID-Release (PHD-SR) examined the factors associating severe COVID-19 among hospitalized pediatric patients. The study found a positive association on severe COVID-19 for pediatric patients with more than 1 underlying medical conditions. It was unclear whether gender plays an important role on COVID-19 infections and severity in pediatrics, but the study reported that severe outcomes were higher for males than females across age groups. Among those with underlying medical conditions, males were more likely to be hospitalized than females on the COVID-19 severity in pediatrics.

Further, multiple studies ([21], [11] and [8]) have stressed out the prevalence of missing race and ethnicity data across age groups. Therefore, this creates challenges in performing statistical analysis and comparisons on severe outcomes are not fully understood. Multiple imputation methods have been proposed to dramatically reduce the bias in statistical power for describing health disparities in CDC COVID-19 case-level surveillance data [21].

Our exploratory analyses regarding descriptive statistics on all levels of severe outcomes (hospitalizations, ICU admissions and deaths) were similar to [18]. However, our study is unique in that we have excluded adults age groups (> 19 years old) but included race and ethnicity factor into our summaries. Our study is also similar to [14] where we presented incidence rate ratios to study the associations between factors affecting severity of COVID-19 in our results section. However, our models are unique in that we fit Bayesian Poisson models rather than multivariate generalized linear models performed by Kompaniyets et al. [12].

3 Data and methods

For this project, we examined the COVID-19 Case Surveillance Public Use Data containing individual level patient records collected by jurisdictions across US states and reported voluntarily to CDC. The available dataset contains 12 variables for all COVID-19 cases ($n = 13,415,836$) from January 1 to December 16, 2020. The 12 variables included reported times, demographics, exposure history, disease severity indicators and outcomes, and presence of underlying medical conditions of de-indentified patients.

As our interest lies in understanding the factors that contribute to severity of infections among pediatric/children patients with COVID-19 in the US, we retrieved only 0-9 and 10-19 age groups observations and aggregated both age groups into 0-19 years. Definition of child varied by state [16], but here, we adopted the convention that children age ranges between 0 to 19 years.

To further ease our computational load and place greater importance on days which had higher COVID-19 cases, we subset our observations beginning from March 1 2020 as this was the time where the official surge in COVID-19 cases in the US had began [15]. We had used the laboratory-confirmed COVID-19 cases, aligning with other studies (e.g. [11] and [18]). After exclusions, 1,754,936 (13.08%) cases were being analyzed.

Patterns of missing data

Upon inspection of time variables, we found missing time data in "cdc_report_dt"[1] (10.98%), "pos_spec_dt" [2] (69.22%) and "onset_dt" [3] (53.10%) respectively. We have decided to exclude "cdc_report_dt" from our analyses because it was deemed deprecated [4]. Further, we found out there were complications in utilizing "pos_spec_dt" and "onset_dt" since the recorded times were inconsistent and missing. These were known to be recorded in the case report form (CRF), leaving many unanswered data. Fortunately, "cdc_case_earliest_dt"[4], did not have any missing entries. This variable was also recommended by CDC [4] for time-related analyses usage. Thus, we only utilized "cdc_case_earliest_dt" for the remaining of our analyses.

Apart from time variables, race and ethnicity factor recorded 4.46% "Missing" and 39.10% of "Unknown" entries. Sex factor recorded 0.29% of "Missing" and 0.99% of "Unknown" entries. Since we have excluded age groups higher than 19 years old, we have also excluded the "Unknown" age group behind. Utility summaries informed us that values under variables coded with "Unknown" were under the specifications submitted to CDC by the US jurisdictions that the value is unknown, while "Missing" implied that jurisdictions did not provide any value [4]. To derive results from the dataset, we utilized only the complete cases for demographic variables (race and ethnicity and sex) and fitted models based on available hospitalization, ICU and death cases.

In disease severity variables, there were only 42.09% hospitalization, 8.84% ICU, 50.17% death, and 12.77% underlying medical condition indicator entries that were complete (not "Missing" and "Unknown"). As underlying medical condition indicator (12.77% available) was substantially under-represented in the dataset, we would like to investigate whether our conclusions change depending on the inclusion of "Missing" and "Unknown" factor levels to the factor. Upon additional exclusions, we finalized to analyzing 983,746 (7.33%) cases.

Exploratory data analysis

Our initial exploratory result was shown in Table (1). There were marginally more female (51.22%) children patients than male (48.78%). However, male generally had higher rates of hospitalizations (1.71% versus 1.69%) and ICU admissions (0.29% versus 0.23%) than female, with only deaths reported at a slightly lower rate (0.05% versus 0.06%) compared to female. When it comes to patients who had underlying health conditions, female reported higher hospitalization rates (3.52% verses 3.47%) but lower ICU admission rates (0.52% verses 0.62%) compared to male. Among those with no underlying conditions, more male hospitalizations (1.02% versus 0.99%) and ICU admissions (0.62% verses 0.52%) compared to female. Negligible differences in death rates were seen between male and female regardless of underlying medical condition indicators.

Moving on to race and ethnicity, White contributed to most of the COVID-19 cases (52.84%), followed by Hispanic/Latino (26.22%), Black (11.89%), Other (4.84%), Asian (2.78%), AI/AN (0.72%) and Hawaiian/PI (0.72%). However, among all race and ethnicity, Black and Hispanic/Latino were consistently among the highest hospitalizations (3.11% and 2.44%), ICU admissions (0.57% and 0.37%), and death rates (0.05% and 0.07%) compared to other races and ethnicity. The hospitalization rate was staggeringly higher among Hispanic/Latino and Black patients with underlying medical conditions, contributing to 5.44% and 5.08% respectively. Despite representing the most cases (100602 or 52.84% of total race and ethnicity), White represents one of the lowest reported hospitalizations, ICU admissions and deaths regardless of underlying health status.

[1] Initial date case reported to CDC.

[2] Initial date of positive COVID-19 specimen collection.

[3] COVID-19 associated symptom onset date (if symptomatic).

[4] Earliest available date reported for CDC.

Table 1: Reported laboratory-confirmed COVID-19 children patients who were hospitalized, admitted to ICU or died stratified by sex and race and ethnicity based on underlying medical condition indicators.

Characteristics	Reported Hospitalizations			Reported ICU admissions			Reported Deaths		
	Among all patients	Among patients with underlying medical condition	Among patients with no underlying medical condition	Among all patients	Among patients with underlying medical condition	Among patients with no underlying medical condition	Among all patients	Among patients with underlying medical condition	Among patients with no underlying medical condition
Sex									
Female (51.22%)	1652/97529 (1.69%)	956/27190 (3.52%)	696/70339 (0.99%)	225/97529 (0.23%)	141/27190 (0.52%)	84/70339 (0.12%)	56/97529 (0.06%)	30/27190 (0.11%)	26/70339 (0.04%)
Male (48.78%)	1591/92878 (1.71%)	911/26220 (3.47%)	680/66658 (1.02%)	272/92878 (0.29%)	162/26220 (0.62%)	110/66658 (0.17%)	49/92878 (0.05%)	28/26220 (0.11%)	21/66658 (0.03%)
Race and Ethnicity									
Hispanic/Latino (26.22%)	1220/49932 (2.44%)	742/13652 (5.44%)	478/36280 (1.32%)	184/49932 (0.37%)	108/13652 (0.79%)	76/36280 (0.21%)	24/49932 (0.05%)	18/13652 (0.13%)	6/36280 (0.02%)
AI/AN (0.72%)	15/1366 (1.1%)	7/286 (2.45%)	8/1080 (0.74%)	2/1366 (0.15%)	1/286 (0.35%)	1/1080 (0.09%)	Nil	Nil	Nil
Asian (2.78%)	94/5287 (1.78%)	43/968 (4.44%)	51/4319 (1.18%)	14/5287 (0.26%)	5/968 (0.52%)	9/4319 (0.21%)	2/5287 (0.04%)	1/968 (0.1%)	1/4319 (0.02%)
Black (11.89%)	705/22634 (3.11%)	454/8935 (5.08%)	251/13699 (1.83%)	128/22634 (0.57%)	87/8935 (0.97%)	41/13699 (0.3%)	23/22634 (0.1%)	19/8935 (0.21%)	4/13699 (0.03%)
Multiple/Other (4.84%)	191/9222 (2.07%)	114/2194 (5.2%)	77/7028 (1.1%)	30/9222 (0.33%)	20/2194 (0.91%)	10/7028 (0.14%)	6/9222 (0.07%)	4/2194 (0.18%)	2/7028 (0.03%)
Hawaiian/PI (0.72%)	17/1364 (1.25%)	13/797 (1.63%)	4/567 (0.71%)	1/1364 (0.07%)	1/797 (0.13%)	Nil	1/1364 (0.07%)	Nil	1/567 (0.18%)
White (52.84%)	1001/100602 (1%)	494/26578 (1.86%)	507/74024 (0.68%)	138/100602 (0.14%)	81/26578 (0.3%)	57/74024 (0.08%)	49/100602 (0.05%)	16/26578 (0.06%)	33/74024 (0.04%)

Abbreviation or meaning: AI/AN = American Indian/Alaska Native, PI = Pacific Islander, Nil = No reported cases.
[a] "Unknown", "Missing" and "Other" categories for sex, hospitalizations, ICU admissions, and deaths were excluded from the reporting table.
[b] Only "Unknown" and "Missing" were excluded from race and ethnicity as "Other" represents more than one race.

Statistical methods and modelling

Firstly, we pursued to investigate the effect of race and ethnicity on the severity of COVID-19 (Hospitalization, ICU admission, and Death). We have fitted the Bayesian Poisson model with log link function to model weekly incidence cases for all severity outcomes separately in COVID-19 children patients (Model 1). To model the heterogeneity in time, we have included random time effects for all observations in each severity, and we kept race as fixed effects. We fixed "White" as the baseline/reference level.

Apart from race and ethnicity, our subsequent modelling (Model 2) included other explanatory variables (sex and underlying medical conditions) to understand how severity of COVID-19 depends on our variables. We fixed "White", "Woman", and "No" as our baseline/reference race, sex, and underlying medical condition indicator respectively. Parameter priors introduced were consistent with Model 1.

The number of weekly incidence (Hospitalization/ICU admission/Death) cases in children, Y_i was assumed to follow Poisson distribution, i.e.,

$$
\begin{aligned}
Y_i &\sim Poisson(\lambda_i) \\
log(\lambda_i) &= \beta_0 + \beta_{TIME,i} + \beta_{RACE[i]} + \log P_i && \text{(Model 1)} \\
log(\lambda_i) &= \beta_0 + \beta_{TIME,i} + \beta_{RACE[i]} + \beta_{SEX[i]} + \beta_{MEDCOND[i]} + \log P_i && \text{(Model 2)}
\end{aligned}
$$

Parameter priors:

$$
\begin{aligned}
\beta_0 &\sim \text{Normal}(0, 10^2) && \text{(Baseline)} \\
\beta_{RACE[i]}, \beta_{SEX[i]}, \beta_{MEDCOND[i]} &\sim \text{Normal}(0, 1^2) && \text{(Fixed race, sex, medical condition effects)} \\
\beta_{TIME,i} &\sim \text{Normal}(0, \sigma^2_{TIME}) && \text{(Random time effects)} \\
\sigma_{TIME} &\sim \text{Half-Cauchy}(0, 1)
\end{aligned}
$$

where each observation index, i has its own time parameter, $\beta_{TIME,i}$. $RACE[i], SEX[i]$, and $MEDCOND[i]$ maps each i to its appropriate category in race, sex and underlying medical condition indicator respectively. P_i is the offset and it represents the number of patients who are at risk of developing severe outcomes (Hospitalization/ICU admisson/Death) with respect to each race for Model 1, each race, sex

and medical condition for Model 2. As we assumed no prior knowledge on the data, non-informative priors were assigned to the parameters.

To implement and perform Bayesian inference to our described models, we utilized the "rstan" package [17]. This allowed us to perform Hamiltonian Monte Carlo (HMC) sampling to estimate our model parameters. For every model, we ran 2 HMC chains with 10000 iterations using 500 burn-in iterations per chain. Model convergence and mixing were assessed appropriately each time. This included examining the smallest effective samples sizes, N_{eff} and largest potential scale-reduction statistics, $\hat{R}$ numerically. A common heuristic would require $N_{eff} > 1000$ so that sufficient sample sizes can be utilized for constructing 95% credible intervals, and $\sqrt{\hat{R}} \leq 1.1$ for HMC chain convergence [2]. Additionally, visual inspection of trace plots for the smallest N_{eff} parameter were conducted.

Posterior estimates excluding burn-in iterations were extracted to compute the posterior medians and 95% credible intervals for all observations. To facilitate comparison of our results, we derived the weekly cumulative incidence cases for all races and ethnicity under Model 1 (additionally by sex and medical condition under Model 2). Besides, to account for differences in laboratory-confirmed COVID-19 patients in each race and ethnicity group (additionally by sex and medical condition under Model 2), we have adjusted our estimated posterior incidence cases for each group by dividing those with the total patient counts in each group respectively. The adjusted posterior cumulative incidence cases by group proportions per 100,000 COVID-19 children patients is:

$$\text{Adjusted cumulative incidence cases} = \frac{Y_{cumulative,i}}{P_i} \times 100,000 \qquad \text{(Equation 1)}$$

where P_i refers to the same offset in the models 1 and 2.

In addition, posterior estimates were used to compute the incidence rate ratios for hospitalization, ICU, and death to compare between races on all severity outcomes among the COVID-19 children patients, taking the reference group as the baseline. The general formula for incidence rate ratio is:

$$\text{Incidence rate ratio} = \frac{\exp(\beta_{target})}{\exp(\beta_{ref})} \qquad \text{(Equation 2)}$$

where β_{target} contains the parameters for the group that we wish to compare against the reference group containing parameters, β_{ref}. For example, to compare the incidence rate ratio between White children (reference group) and Black children, the incidence rate ratio is $\exp(\beta_{RACE[i]})$, where $RACE[i]$ corresponds to a Black observation. This is because we have treated β_0 as the baseline coefficient for White, and hence, $\beta_{RACE[i]}$ is the only remaining coefficient that makes the difference after cancellation in the numerator and denominator of (Equation 2).

As underlying medical condition constituted most "Missing" and "Unknown" entries (87.23%) in comparison to sex and race and ethnicity, we included and excluded these entries in separate model fits with Model 2 and assess how severity changes depending on our variables as part of our multiverse analysis. Incidence rate ratios using the corresponding posterior estimates were compared across race and ethnicity, sexes and underlying medical conditions.

4 Results

Model diagnostics were performed to inspect whether HMC samples converges and mixes well for all runs in Model 1. All parameters for hospitalization, ICU, and death runs have shown $\hat{R}$ were close to 1 indicating chains have been well mixed and have all sampled the target distribution. Although during hospitalization run, $N_{eff} = 549 (< 1000)$ was obtained, trace plots indicated good mixing and convergence. All trace plots with smallest N_{eff} were presented in Appendix (A).

In Figure (1), our posterior median estimates under Model 1 suggested stark disparities in race among children COVID-19 patients. In all severity levels, we observed that Black followed by Hispanic/Latino children patients consistently experienced higher rates in cumulative adjusted cases compared to other races. Widening gaps were noticeably seen between Black and other races along the rest of 2020 especially in terms of ICU admission and Death. The adjusted cumulative cases surpassed more than 1750 for hospitalization, 175 for ICU, and reaching almost 40 deaths per 100,000 COVID-19 children patients for

Black. In contrast, we observe that White children were among the lowest adjusted cumulative cases per 100,000 COVID-19 patients, although death cases seemed to be increasing wildly beyond week 40 in 2020.

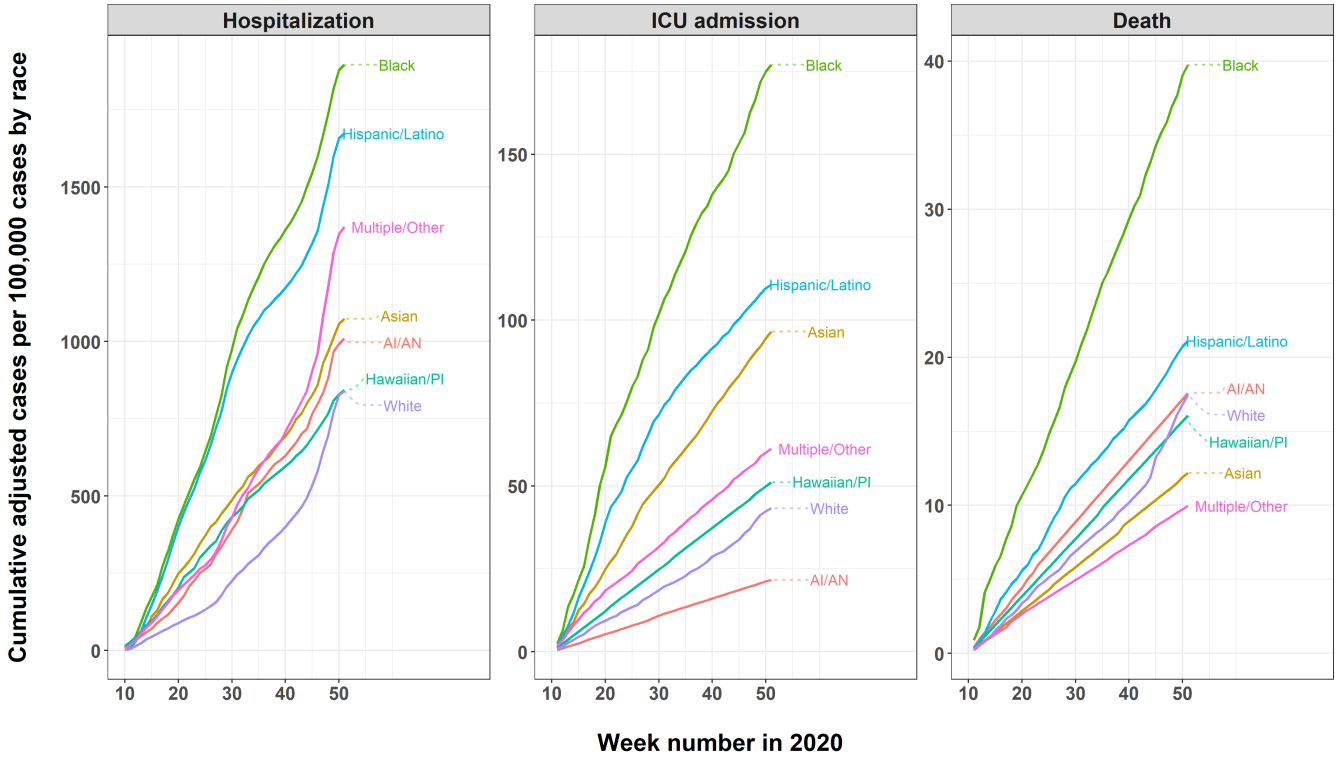

Figure 1: Estimated posterior median adjusted cumulative cases per 100,000 laboratory-confirmed COVID-19 children patients by race in all severity levels: Hospitalization, ICU admission and death since week 10 (first week of March in 2020) according to Model 1.

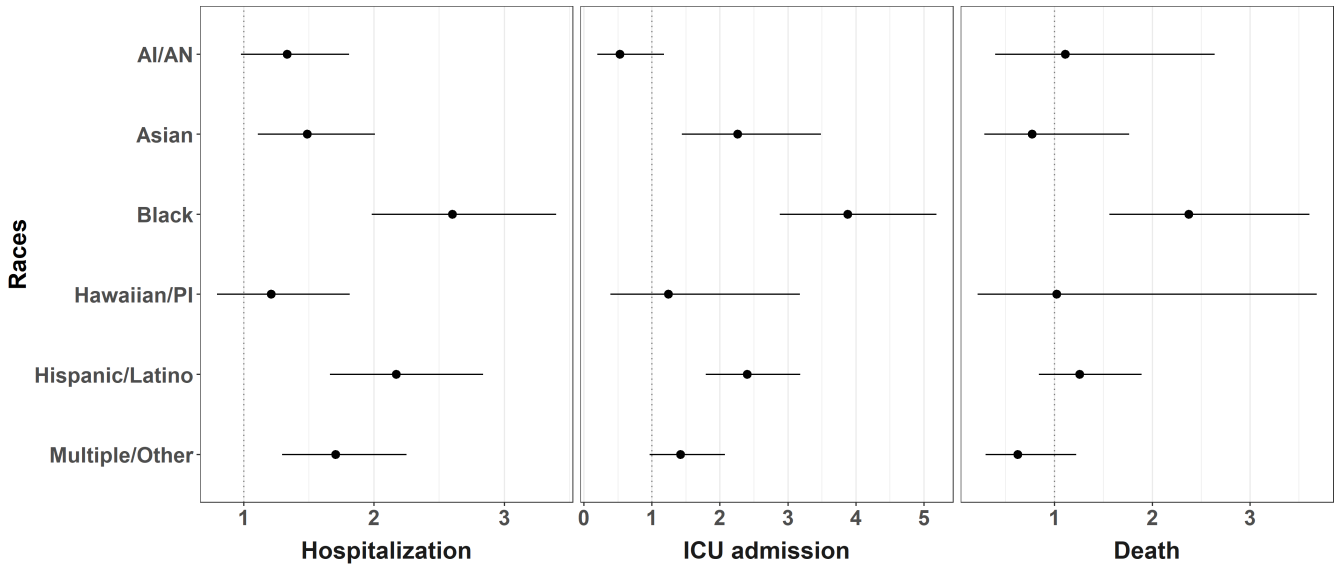

Figure 2: Plot of estimated incidence rate ratios (median and 95% credible intervals) in all severity levels: Hospitalization, ICU and death of laboratory-confirmed COVID-19 children patients by race under Model 1. Dotted vertical dashed line at 1 indicates the relative rate for comparison group (White). Rates > 1 indicates higher incidence rate compared to White and vice versa.

Incidence rate ratios plot results were presented in Figure (2). Black children reported two to three-fold of hospitalizations (Median= 2.60; 95% Credible Interval= $[1.98, 3.3]$) followed by Latino $(2.17; [1.66, 2.84])$, Multiple/Other $(1.70; [1.29, 2.25])$, Asian $(1.49; [1.11, 2.01])$, AI/AN $(1.33; [0.98, 1.81])$, and Hawaiian/PI $(1.21; [0.79, 1.81])$ compared to White. Our findings were consistent with Figure (1). For ICU, only the

rate ratios for AI/AN $(0.53, [0.20, 1.18])$ and Hawaiian/PI $(1.24, [0.39, 3.18])$ were similar to Whites, while Black $(3.88, [2.88, 5.18])$, Hispanic/Latino $(2.40, [1.79, 3.18])$ and Asian $(2.26, [1.44, 3.48])$ continues to be the top three highest rates compared to White. For death, rate ratios for all races were comparable with White except for Black $(2.37, [1.56, 3.62])$.

Next, we applied Model 2 (including sex, and underlying medical condition factors). "Missing" and "Unknown" observations in underlying medical condition factor were first excluded. Model diagnostics revealed uncertainty in estimates as convergence and mixing results were unsatisfactory, especially in ICU and Death runs ($N_{eff} = 486$ and $N_{eff} = 110$). Perhaps, the runs had been caused by lesser incidence cases feeding into our model, making accurate posterior estimates difficult. Upon examination of the incidence rate ratios for race, we did not find significant difference to Figure (2). Hence, results were omitted. With regards to sexes, we observed that male children patients have significantly higher ICU admissions than female, but no significant differences in hospitalizations and deaths. Also, COVID-19 children patients with underlying medical conditions have about 3 times more likely to be hospitalized, admit to ICU and die.

Finally, "Missing" and "Unknown" observations for underlying medical condition were included to Model 2. Firstly, model diagnostics tells us that the smallest N_{eff} was 3789 for hospitalization, 3854 for ICU, but only 45 for death. Coupled with trace plots, HMC chains for death indicated significant poor convergence and mixing. Combining evidence from previous runs, it was likely to be caused by the low incidence cases in death as suggested in Table (1) resulting in poorer estimation ability for our Bayesian model. Nevertheless, discussions were made with regards to the available results.

Examining the incidence rate ratios for race again suggest no pronounced difference in credible interval estimates compared to Figure (2). Exact incidence rate ratio statistics were presented in Table (2). For that reason, we hypothesized that sex and underlying medical factors were not likely to be correlated with race and ethnicity when it comes to children COVID-19 patients. According to Table (2) and Figure (3), we noted lower ICU admission and death rate ratio estimates in "Missing" (ICU admission: $0.24; [0.19, 0.30]$; Death: $0.44; [0.31, 0.63]$) and "Unknown" (ICU admission: $(0.66; [0.48, 0.89]$; Death $0.38; (0.18, 0.75))$ underlying medical conditions, compared to patients with no underlying medical conditions. However, "Unknown" has higher hospitalization rates $(1.66; [1.44, 1.90])$. This suggested to us that underlying medical condition data may not be missing at random.

Table 2: Estimated incidence rate ratios of laboratory-confirmed COVID-19 children patients in all severity levels: Hospitalization, ICU admission and death since week 10 of 2020 under Model 2.

	Incidence rate ratio (95% credible interval)		
Characteristics	**Hospitalization**	**ICU admission**	**Death**
Race and Ethnicity			
AI/AN [1]	1.50 (1.21,1.86)	0.79 (0.30,1.76)	1.32 (0.47,3.05)
Asian [1]	1.60 (1.35,1.89)	2.55 (1.69,3.73)	0.81 (0.31,1.78)
Black [1]	2.48 (2.19,2.81)	3.66 (2.93,4.58)	2.18 (1.52,3.09)
Hawaiian/PI [1]	1.00 (0.71,1.38)	0.84 (0.28,2.04)	0.82 (0.18,2.81)
Hispanic/Latino [1]	2.14 (1.89,2.42)	2.47 (2.00,3.05)	1.24 (0.88,1.72)
Multiple/Other [1]	1.86 (1.61,2.15)	1.83 (1.31,2.55)	0.70 (0.34,1.33)
Sex			
Male [2]	0.99 (0.91,1.07)	1.34 (1.14,1.58)	1.10 (0.84,1.44)
Underlying medical condition indicator			
Missing [3]	0.92 (0.81,1.03)	0.24 (0.19,0.30)	0.44 (0.31,0.63)
Unknown [3]	1.66 (1.44,1.90)	0.66 (0.48,0.89)	0.38 (0.18,0.75)
Yes [3]	3.30 (2.90,3.75)	3.75 (3.03,4.63)	2.99 (2.00,4.41)

[1] The reference group is "White" children patients.

[2] The reference group is "Female" children patients.

[3] The reference group is children patients with "No" underlying medical condition.

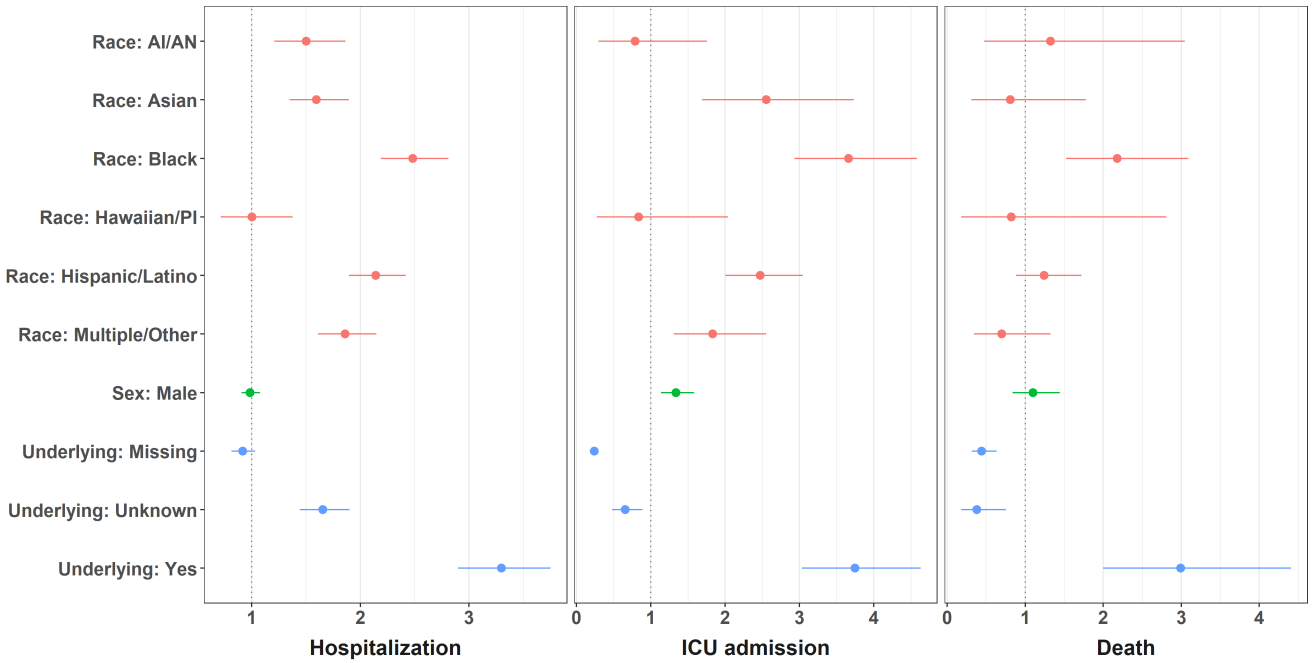

Figure 3: Plot of estimated incidence rate ratios (median and 95% credible intervals) in all severity levels: Hospitalization, ICU and death of laboratory-confirmed COVID-19 children patients by race, sex and all underlying medical condition factors with Model 2. Dotted vertical dashed line at 1 indicates the relative rate for the comparison group (White race, Female sex, No underlying medical condition).

However, our trained models may be biased by the fact that additional factors that were not considered to modelling. One obvious one is accounting for the difference across children age groups, which may be different for 0-9 years and 10-19 years old from the dataset. Study by Preston et al. [14] shows that older children are more likely to develop severe COVID-19 illness compared to younger children. This is something that we did not account for, and our results may suggest otherwise.

5 Conclusion and discussion

Overall, our analyses revealed the main factors that affected the severity of COVID-19 infections among children patients in the US using the CDC COVID-19 Case Surveillance public use dataset. Our exploratory analysis and model results both suggested that the severity of COVID-19 infections did depend on children patient's race and ethnicity, and all race groups have proportionally higher hospitalization rates than White race. Black and Hispanic/Latino children patients were more double to triple more likely to ended up hospitalized and admitted to ICU compared to White children patients. Moreover, our models suggested children patients who possessed underlying medical conditions were three-fold more likely to be hospitalized, admitted to ICU and die as well. While no statistically significant differences on hospitalization and mortality rates in both sexes, male were more likely to be admitted ICU than female. Overall, our findings provided supporting evidence to many prior studies (e.g.[11], [1] and [3]).

However, our study is subject to many limitations. First, serious considerations must be taken for interpretation of our model estimates as incidence cases for all severity outcomes fell short when considering the children age group category (0-19 years) compared to older age groups. All reports and outcomes were provisional and might change accordingly during the surveillance period [4]. With such small number of cases especially in death counts for children patients, coupled with utilizing only data from early stages of pandemic, it can be significantly difficult to quantify with certainty on the association of factors on the severity of COVID-19 outcomes in the children age group. Furthermore, although reports were made using only laboratory-confirmed cases, doubts on whether a clear distinction should be made for probable cases to assess the severity of COVID-19 outcomes.

Next, our model estimates were significantly dampened by the large number of missing entries in the dataset across all variables. This likely resulted in under-estimates of the true prevalence in the variables. Also, smaller age group sizes, for example, a 5 years gap of age group data, 0-5 years, 5-11 years and so on,

were not prepared due to privacy concerns [13]. Overall, the dataset is most likely to be under-reported and it represents only a subset of all COVID-19 cases in the US because capturing each and every community cases would arguably impossible.

Despite the limitations, our findings contributed in understanding factors that severity of COVID-19 infections in US children. Our evidence shown that unfortunately, race and ethnicity disparity issues have to accounted even for young children. Given the persistence of racialized injustice in the US across age groups, this reflects the urgent need to sustain the under-represented communities through culturally responsive public health care efforts in the US. Besides, public health practitioners should consider the potential need for clinically managing COVID-19 infected children with underlying medical conditions to minimize their risk of severe outcomes.

Given more time and data, we would like to make justifiable conclusions on the relationship between possible factors on severity of COVID-19 outcomes among children. Additional socioeconomic factors such as family educational level and income are worth exploring. Obtaining data across geographical locations using CDC's Case Surveillance Public Use Data with Geography may open doors to deepen our understanding on the disparities in COVID-19 outcomes among children in US. We also wish to explore different types of models and perform sensitivity analysis to assess the robustness of our findings.

A

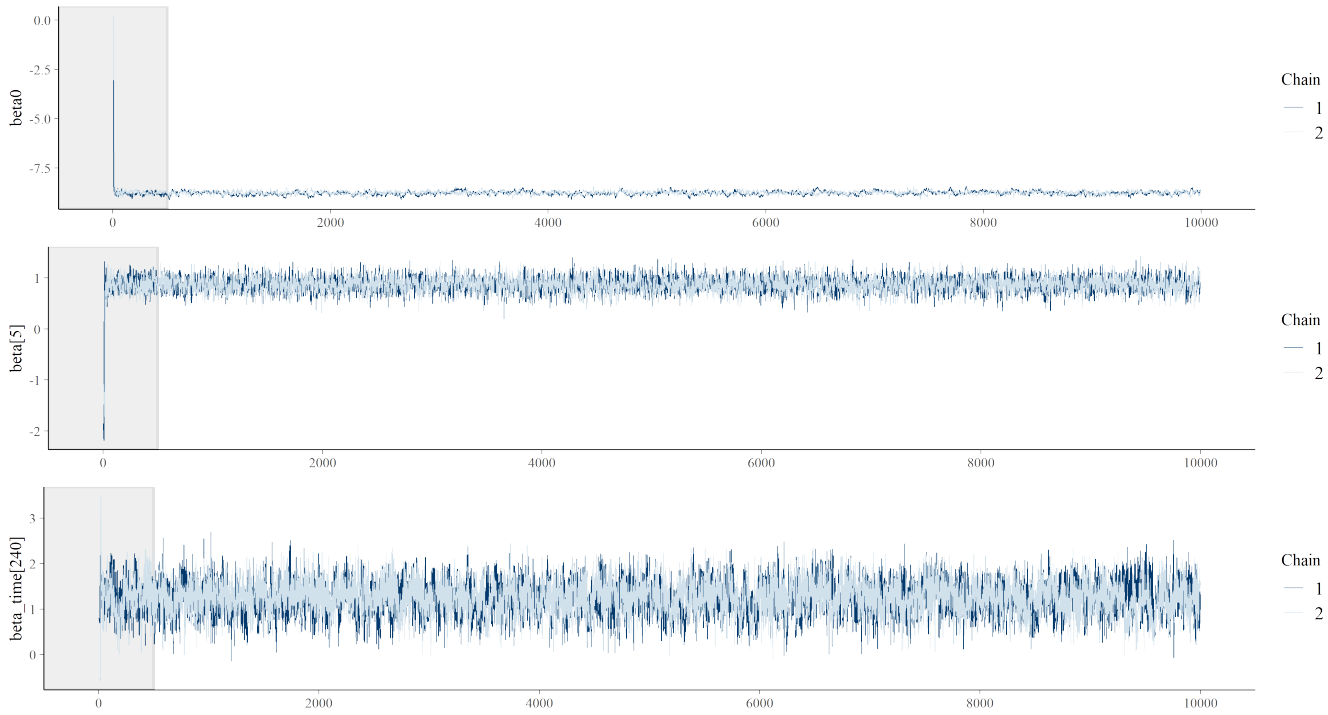

Figure 4: Trace plots of the parameters with the smallest N_{eff} during the generation of HMC chains for hospitalization, ICU admission, and death (top to bottom panel) respectively under (Model 1).

References

[1] D. Bixler. Sars-cov-2–associated deaths among persons aged 21 years—united states, february 12–july 31, 2020. *MMWR. Morbidity and mortality weekly report*, 69, 2020.

[2] S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.

[3] R. Castagnoli, M. Votto, A. Licari, I. Brambilla, R. Bruno, S. Perlini, F. Rovida, F. Baldanti, and G. L. Marseglia. Severe acute respiratory syndrome coronavirus 2 (sars-cov-2) infection in children and adolescents: a systematic review. *JAMA pediatrics*, 174(9):882–889, 2020.

[4] C. COVID. Case surveillance public use data profile. 19.

[5] F. S. Dawood, C. A. Porucznik, V. Veguilla, J. B. Stanford, J. Duque, M. A. Rolfes, A. Dixon, P. Thind, E. Hacker, M. J. E. Castro, et al. Incidence rates, household infection risk, and clinical characteristics of sars-cov-2 infection among children and adults in utah and new york city, new york. *JAMA pediatrics*, 176(1):59–67, 2022.

[6] C. for Disease Control, Prevention, et al. Cdc covid data tracker: demographic trends of covid-19 cases and deaths in the us reported to cdc, 2020.

[7] C. for Disease Control, Prevention, et al. Covid-net: a weekly summary of us covid-19 hospitalization data. *Accessed August*, 27, 2020.

[8] J. A. Gold, J. DeCuir, J. P. Coyle, L. M. Duca, J. Adjemian, K. N. Anderson, B. N. Baack, A. Bhattarai, D. Dee, T. M. Durant, et al. Covid-19 case surveillance: Trends in person-level case data completeness, united states, april 5–september 30, 2020. *Public Health Reports*, 136(4):466–474, 2021.

[9] L. R.-G. C. A. C. G. E. O.-O. J. H. B. M. D. B. Hannah Ritchie, Edouard Mathieu and M. Roser. Coronavirus pandemic (covid-19). *Our World in Data*, 2020. https://ourworldindata.org/coronavirus.

[10] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The lancet*, 395(10223): 497–506, 2020.

[11] L. Kim, M. Whitaker, A. O'Halloran, A. Kambhampati, S. J. Chai, A. Reingold, I. Armistead, B. Kawasaki, J. Meek, K. Yousey-Hindes, et al. Hospitalization rates and characteristics of children aged¡ 18 years hospitalized with laboratory-confirmed covid-19—covid-net, 14 states, march 1–july 25, 2020. *Morbidity and Mortality Weekly Report*, 69(32):1081, 2020.

[12] L. Kompaniyets, N. T. Agathis, J. M. Nelson, L. E. Preston, J. Y. Ko, B. Belay, A. F. Pennington, M. L. Danielson, C. L. DeSisto, J. R. Chevinsky, et al. Underlying medical conditions associated with severe covid-19 illness among children. *JAMA network open*, 4(6):e2111182–e2111182, 2021.

[13] B. Lee, B. Dupervil, N. P. Deputy, W. Duck, S. Soroka, L. Bottichio, B. Silk, J. Price, P. Sweeney, J. Fuld, et al. Protecting privacy and transforming covid-19 case surveillance datasets for public use. *Public Health Reports*, 136(5):554–561, 2021.

[14] L. E. Preston, J. R. Chevinsky, L. Kompaniyets, A. M. Lavery, A. Kimball, T. K. Boehmer, and A. B. Goodman. Characteristics and disease severity of us children and adolescents diagnosed with covid-19. *JAMA network open*, 4(4):e215298–e215298, 2021.

[15] J. M. Shultz, A. Perlin, R. G. Saltzman, Z. Espinel, and S. Galea. Pandemic march: 2019 coronavirus disease's first wave circumnavigates the globe. *Disaster medicine and public health preparedness*, 14 (5):e28–e32, 2020.

[16] B. Sisk. Children and covid-19: State data report-a joint report from the american academy of pediatrics and the children's hospital association. 2020.

[17] Stan Development Team. RStan: the R interface to Stan, 2020. URL `http://mc-stan.org/`. R package version 2.21.2.

[18] E. K. Stokes, L. D. Zambrano, K. N. Anderson, E. P. Marder, K. M. Raz, S. E. B. Felix, Y. Tie, and K. E. Fullerton. Coronavirus disease 2019 case surveillance—united states, january 22–may 30, 2020. *Morbidity and Mortality Weekly Report*, 69(24):759, 2020.

[19] H. Thomson. Children with long covid, 2021.

[20] Z. Wu and J. M. McGoogan. Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention. *jama*, 323(13):1239–1242, 2020.

[21] G. Zhang, C. E. Rose, Y. Zhang, R. Li, F. C. Lee, G. Massetti, and L. E. Adams. Multiple imputation of missing race and ethnicity in cdc covid-19 case-level surveillance data. *International Journal of Statistics in Medical Research*, 11:1–11, 2022.