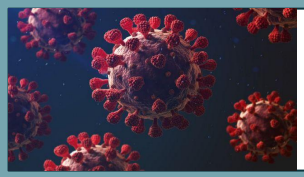

Correlation of Certain Neonatal Abnormalities with NICU Admission Before and During COVID-19

Jane Chen, Chelsea Ji, Soo Lee, Malika Top

Background:

1. Past research suggests there was a decrease in the total number of NICU admissions from 2019 to 2021(Rasmussen, 2021).
2. Given the prevalence of the COVID-19 pandemic over the past two years, we wanted to explore the impacts that the presence of COVID-19 had on mothers and newborns.
3. We used neonatal statistics provided by the Vital Statistics Online Data Portal from the Center of Disease Control. The Natality-related variables we used are NICU admission rates and birth abnormalities (whether the neonatal required immediate ventilation, ventilation for >6 hours, surfactant, antibiotics, or experienced a seizure). Surfactant is a medication (made of fats and proteins) used to treat respiratory distress syndrome (RDS) in newborns where it coats the tiny air sacs in the lungs and helps to keep them from collapsing by preventing alveoli from sticking together when the newborn exhales.
4. We looked at the years of 2019 and 2021, as 2019 represents the period before the onset of COVID-19 and 2021 represents the period after the initial onset and spread of COVID-19.

Research Questions:

Is the COVID-19 pandemic correlated with NICU admissions in the United States? Which neonatal abnormalities are most highly correlated with NICU admits in the United States?

Data Preparation & Methodology:

1. We began with a .txt file for each year from the Vital Statistics Online Data Portal, where variables were broken into columns that were separated by blanks and each row was a birth that occurred in that year. For 2019 and 2021 each, we created 20 data frames where we read in 50,000 rows at a time into a table and randomly selected 5,000 from that data frame into a separate one.
2. After, we combined all the subsetted data frames, repeating for both years. We located our congenital abnormality variables of interest in the 50th column where the information was grouped together into one string. An issue arose with the read.table() command misreading data, so for certain rows, that information was in the 49th column with random strings of 1s and 0s in our intended column of interest. Thus, we manually shifted the correct column values into the appropriate place.
3. For missing data, we filled them in with a string placeholder in order to split the string and get the individual column value for each abnormality (Y - yes, N - no, U - unknown, Z - NA values). This is appropriate because given the enormity of CDC's data set on births and "unknown" being an option in the questionnaire, we assumed that the missing data resulted from data management issues, which would make the missing data missing completely at random (MCAR). We were only interested in the instances of Ys and Ns, so we created another data frame that included rows with just these results and then added columns for the numerical version of the outcomes (Y=1, N=0) for each abnormality.
4. In total we sampled 100,000 rows from each year, out of 3,757,582 and 3,669,928 rows from 2019 and 2021 respectively. After removing rows with missing values, we were left with a data frame of 169,767 rows which is what our two sample t-tests and classification tree ran on.

Data & Visualizations:

```
       202101    11403          11
1                          29094     1
101101  0  011 X1  16 1
320501101    00116 1      000000  1  1
88888    88888    88888       0211          09  061
N10000000000001111N1                61121.72     1151
145 13031          NNNNNN111111NXXX11N00 111
NNNNN111111      NN 11
NYYYNN1111110      11X111111      NNNNN 11111 1
1N2211     0731885    1    9                    M 05
2020     34051     340512350 052       YNYNYN
111111 0      NNNNNN111111NNNNNN1111111      NYY1
       202101    16343          11
1                          29094     1
201101  0  011 X1  16 1
320501101    00114 1      010000  2  2
03605    88888    03605       0311          09  061
N10000000000001111N1          68129.63     1951
```

Figure 1: Screenshot of the raw data (.txt file) from 2021.

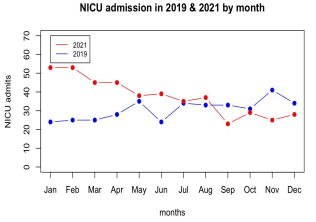

Figure 2: NICU admits based on random sample of 5000 rows taken from each year.

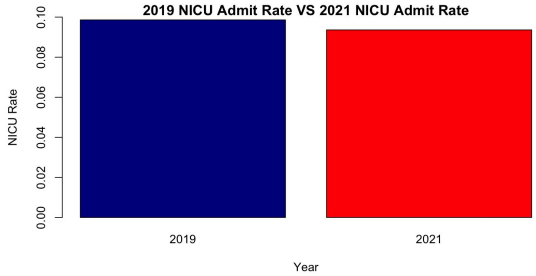

Figure 3: Barplot of NICU admission rates in 2019 compared to 2021 shows a decrease of 5.56% in the total number of NICU admission rate in 2019 vs. 2021

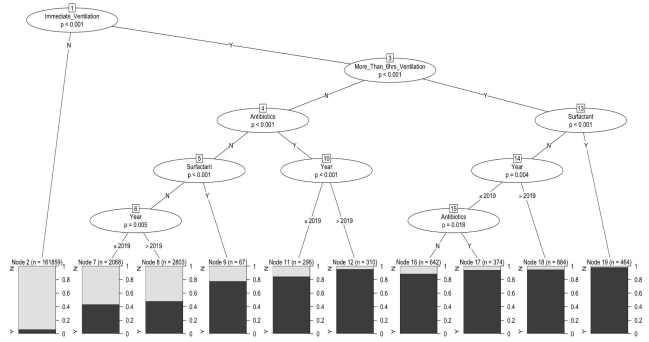

Figure 4: Regression tree based on the following predictors: year (2019/2021), assisted ventilation, assisted ventilation for >6 hours, surfactant administration, antibiotics administration, and observation of seizure. The bar graphs at the bottom respectively show the proportion of likelihood for whether the newborn was admitted into the NICU unit or not.

Results:

Is there a significant difference in abnormalities between 2019 (pre-COVID) and 2021 (during COVID)?
1. Two sample t-test on the NICU admission rate in 2019 and 2021 gives a **two-sided p-value = 0.002269**, which is small enough to rule out the possibility that there is no difference for the mean number of NICU admits across the two years.
2. After running a t-test on the seizure rate in 2019 and 2021, our **two-sided p-value = 0.002021**, which is small enough to rule out the possibility that there is no difference for the mean number of newborns who have seizures across the two years.
3. After running a t-test on the proportion of newborns who required greater than six hours of assisted ventilation in 2019 and 2021, **our two-sided p-value < 2.2e-16** < 0.001, therefore being significantly small enough to rule out the possibility that there is no difference for the mean number of newborns who needed greater than six hours of assisted ventilation across the two years.
4. The barplot suggests that 2019 had a higher rate of NICU admission than 2021. This observation is supported by literature in examining COVID-19 effects on NICU admission as Frontiers in Pediatrics, in a "post-hoc analysis based on data from the 46 NICUs", "found a decrease of 10.3% in the total number of NICU admissions (n = 7,499 in 2020 vs. n = 8,362 in 2019)."

Is there more correlation between abnormalities in 2019 versus 2021?
1. To answer this question, we developed meaningful visualizations for the data by creating heatmaps and regression trees. Based on the regression tree, whether the newborn needed assisted ventilation was the most important predictor of whether the newborn was admitted to NICU for both 2019 and 2021.

Conclusion:

1. Through our study, we determined that 2019 had significantly more NICU admits than 2021. **While we cannot prove that the pandemic itself caused a decrease in babies admitted to the NICU, the fact that there is a significant difference may suggest that there were protocol changes in newborn delivery and care from pre-pandemic to during.** The results of the three two-sample t-tests that we ran provide evidence that whether a baby was born in 2019 or 2021 had an impact on NICU admission, seizure rate, and if they were on assisted ventilation for more than six hours.
2. The regression tree model suggests that the year (2019 versus 2021) does matter, but perhaps not as much as immediate ventilation, which is the most important factor in determining NICU admission for neonates. Assisted ventilation for greater than 6 hours was also more important in determining NICU admission rates than the year. However, for neonates who needed greater than 6 hours of ventilation and who did not need surfactant, the year was significant in determining NICU admission. Additionally, if a neonatal needed immediate ventilation (but not over 6 hours) as well as antibiotics, year was significant in determining NICU admission. Therefore, while the year was significant, it was not the most important factor in determining NICU admission.
3. In the future, our model could be used to investigate predictors for whether a newborn would require NICU admission based on the treatments that they receive after birth. However, there are limitations to our model as a newborn's health consists of more variables than the six that we chose for this study.

References:

Our data source: https://www.cdc.gov/nchs/data_access/VitalStatsOnline.htm
Rasmussen, M. I., Hansen, M. L., Pichler, G., Dempsey, E., Pellicer, A., EL-Khuffash, A., A, S., Piris-Borregas, S., Alsina, M., Cetinkaya, M., Chalak, L., Özkan, H., Baserga, M., Sirc, J., Fuchs, H., Ergenekon, E., Arruza, L., Mathur, A., Stocker, M., … Greisen, G. (2021). Extremely preterm infant admissions within the safeboosc-iii consortium during the covid-19 lockdown. Frontiers in Pediatrics, 9. https://www.frontiersin.org/articles/10.3389/fped.2021.647880