

Modeling of COVID-19 total hospitalizations in the United States

Jonathan Kopel BS, Thomas E. Tenner Jr, PhD, Gregory L. Brower DVM, PhD

ABSTRACT

The SARS-CoV-2 (COVID-19) virus continues to increase across the globe affecting all aspects of modern life. It remains unknown whether COVID-19 hospitalizations can be effectively modeled using regression analysis. Specifically, it is unknown which regression model may accurately reflect past or future trends in COVID-19 hospitalizations. We wanted to see whether we could develop a simple model to describe both previous and future COVID-19 hospitalizations. The graph for total hospital admissions for COVID-19 shows a curve similar to a sine wave with peaks in total hospitalizations occurring in April, July, and December. We used regression analysis for total COVID-19 hospitalizations to provide insight into potential factors influencing COVID-19 hospitalizations and predict future hospitalizations. We found that the total hospitalizations in the United States followed a sine-wave distribution with peaks in hospitalizations every 3.5 months between April and November 2020. However, the sine-wave distribution for COVID-19 disappeared when the model was extended to December 2020. In general, mathematical modeling of hospitalizations works best when there is an established pattern of disease transmission from multiple years of data collection; COVID-19 is a novel virus for which we have less than a year's worth of data from which to draw conclusions. Furthermore, there remains uncertainty about the trajectory of COVID-19 cases and hospitalizations in the future, particularly with the recent emergency use authorization of the Pfizer and Moderna COVID-19 vaccines.

Keywords: SARS-CoV-2, COVID-19, hospitalizations, ventilator, morbidity, and mortality

INTRODUCTION

The SARS-CoV-2 (COVID-19) virus continues to increase across the globe affecting all aspects of modern life. Despite the increasing numbers of SARS-CoV-2 infections, there is increased pressure to re-open economic activity and bring “normalcy” to people's lives. The primary method for detecting and monitoring the spread of the SARS-CoV-2 virus is the reverse-transcriptase–polymerase-chain-reaction (RT-PCR) test and clinical symptoms.¹ However, the RT-PCR requires expensive equipment and trained technicians at certified

laboratories, requires wait times to generate results, and has risks for false-positive results in patients due to excessively high cycle counts.¹ Tracking hospitalizations from the SARS-CoV-2 provides another method for assessing the spread and severity of COVID-19. The number of hospitalizations is a better metric for severe disease (rather than asymptomatic cases) and a better warning indicator that hospital capacity may be overwhelmed.² At the beginning of the pandemic, hospitalizations rose steeply as the pandemic swept through parts of the United States (US), particularly in New York and New Jersey. In response, hospital admissions for other problems, including elective procedures, fell dramatically, leading many hospitals to operate at less than 50 percent capacity.² This also included reduced hospital admission for acute medical disorders, including stroke and acute myocardial infarction.^{3–7}

Corresponding author: Jonathan Kopel
Contact Information: Jonathan.Kopel@tuhsc.edu
DOI: 10.12746/swrccc.v9i37.811

Since this is a new virus, it remains unknown whether COVID-19 hospitalizations can be effectively modeled using regression analysis. Specifically, it is unknown which regression model may accurately reflect past or future trends in COVID-19 hospitalizations. The Centers for Disease Control (CDC) uses several mathematical models to predict national and state numbers of new and total COVID-19 cases, hospitalizations, deaths every four weeks. These forecasts used different types of data (e.g., COVID-19 data, demographic data, mobility data), methods, and estimates of interventions, such as social distancing and use of face coverings.⁸ Given the uncertainty and variability with computer models, the CDC uses an ensemble of different computer forecasts for comparison of what may happen in the near future.⁸ The ensemble is a collaboration among the CDC, 21 academic research groups, five private industry groups, and two government-affiliated groups.⁸ Each of the models used for COVID-19 forecasting provides a median for the predicted distribution and 11 prediction intervals ranging from a 10% prediction interval to a 98% prediction interval.⁸ These forecasts are updated every 4 weeks as more data accumulate to better ensure the predicative ability of the models and account for new errors. We wanted to see whether we could develop a simple model to describe both previous and future COVID-19 hospitalizations. The graph for total hospital admissions for COVID-19 shows a curve similar to a sine wave with peaks in total hospitalizations occurring in April, July, and December. In this paper, we used regression analysis for total COVID-19 hospitalizations to provide insight into potential factors influencing COVID-19 hospitalizations and predict future hospitalizations.

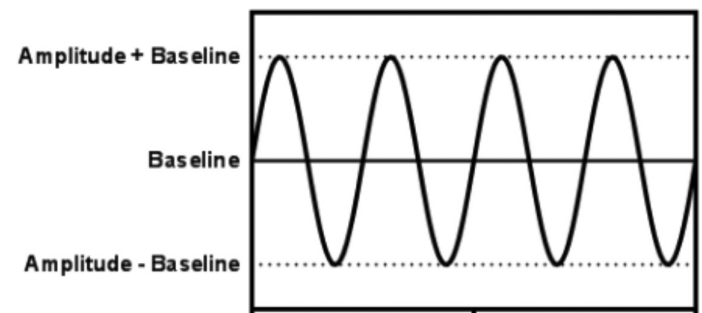
METHODS

Data on total US COVID-19 hospitalizations, positive COVID-19 tests, and COVID-19 mortality were obtained from the CDC and the COVID-19 Tracking Project at the Atlantic (<https://covidtracking.com/data/national>). The initial sine wave model was based on a regression analysis of data from April to November 2020. The model was subsequently reassessed by extending the data through December 2020 when new data became available. The data were analyzed

using the Graph-pad Prism 9 software for data analysis and data presentation.

RESULTS

Using COVID-19 hospitalization data from the COVID-19 Tracking Project at the Atlantic, the total number of COVID-19 hospitalizations was plotted versus time (Figure 1A and B). A regression analysis was performed on Figure 1A and B using a sinewave with a non-zero baseline model as shown below:



The equation used for this regression analysis is given below:

$$y(t) = A \cdot \sin\left(\frac{2\pi t}{T} \pm \Phi\right) + B$$

$y(t)$ – total COVID-19 hospitalizations

t – time in days

T – period in days

A – the amplitude of the waveform

Φ – the phase angle that the waveform has shifted either left or right in radians

B – baseline or average value

Using 209 data points from April to November 2020, the regression analysis showed that the coefficient of determination (r^2) for the COVID-19 hospitalization was 0.8840 (Figure 1A). This model includes the first two peaks in COVID-19 hospitalizations in April and July of 2020. The amplitude (12,963–14,337 hospitalization patients), period (101.4–104.6 days), phase shift (30.77–31.04 days), baseline (41,396–42,446 hospitalized patients), and frequency (0.00956/day–0.00968/day) parameters for the model are shown in Table 1A. Each of these values was reported with a confidence

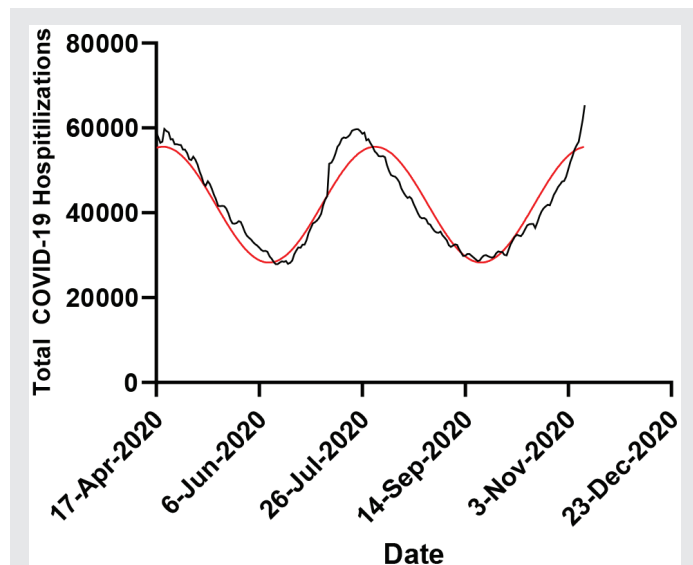


Figure 1A. Total daily COVID-19 hospitalizations from April–November 2020.

interval of 95%. The regression analysis showed that the baseline or average number of hospitalizations (B) was estimated to be between 41,396 to 42,446 COVID-19 patients. The maximum increase or decrease (amplitude– A_{max}) of COVID-19 patients from the baseline was between 12,963–14,337 COVID-19 patients.

A second regression analysis using a sine wave with a non-zero baseline was performed by extending the data through December 2020. Using 256 data points from April to December 2020, the regression analysis showed that the coefficient of determination (r^2) for the COVID-19 hospitalization was 0.1178 (Figure 1B). Compared to the regression model in Figure 1A, this model includes all three peaks (April, July, and December) in COVID-19 hospitalizations in April and July of 2020. The amplitude (7790–15816 hospitalization patients), period (70.05 days–76.29 days), phase shift (5.355–6.858 days), baseline (48,307–53,975 hospitalized patients), and frequency (0.01311/day–0.01428/day) parameters for the model are shown in Table 1B. The regression analysis showed that the baseline or average number of hospitalizations (B) was estimated to be between 48,307–53,975 COVID-19 patients. The maximum increase or decrease (amplitude– A_{max}) of COVID-19 patients from the baseline was between 7,790–15,816 COVID-19 patients.

Table 1A. Sine Wave with Non-Zero Baseline Regression Analysis from April–November 2020

Sine Wave Parameters	
Amplitude (number of hospitalizations)	12963 to 14337*
Wavelength (days)	101.4 to 104.6
Phase Shift (days)	30.77 to 31.04
Baseline (number of hospitalizations)	41396 to 42446
Frequency	0.009561 to 0.009858
Goodness of Fit	
Degrees of Freedom	205
R squared	0.8840
Sum of squares	2589421074
Sy.x	3554
Number of Points	
# of X values	209
# Y values analyzed	209

*95% confidence interval

The first peak in hospitalizations occurred in April and corresponded to the increase in total COVID-19 hospitalizations in certain states, including New York (Figures 1 and 2). After June, the daily increase in COVID-19 hospitalizations decreased until the middle of July, when the COVID-19 hospitalizations peaked a second time. This second peak was due to outbreaks in the South and West, including Texas (Figure 2). Subsequently, the daily COVID-19 hospitalizations began to increase in November. This third peak is ongoing and may be related to colder weather and holiday (Thanksgiving and Christmas) gatherings. Different patterns seen in different regions and states reflect differences in local and state policies for managing the spread of COVID-19 at the beginning and later stages of the COVID-19 pandemic.

Data from the Atlantic COVID-19 Tracking Project show a similar trend in cases and hospitalizations for the Northeast, Midwest, Southern, and Western regions of the US (Figures 2 and 3). As shown in Figure 4, Texas, New York, and California provide examples of

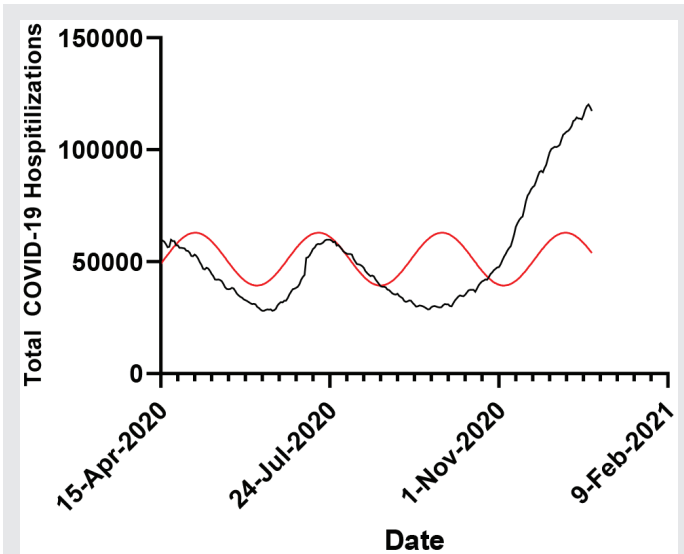


Figure 1B. Total daily COVID-19 hospitalizations from April–December 2020.

the hospitalizations that occurred at different times during the year as COVID-19 continued to spread. The first peak in COVID-19 cases and hospitalizations coincided with the initial outbreak with cases and deaths dominated by New York, New Jersey, Massachusetts, and Connecticut. The second peak coincided with the outbreaks across the South (Texas, Florida, Arizona, etc.). There was little or no peak in the Northeast and

Table 1B. Sine Wave with Non-Zero Baseline Regression Analysis from April–December 2020

Sine Wave Parameters	
Amplitude (number of hospitalizations)	7790 to 15816*
Wavelength (days)	70.05 to 76.29
Phase Shift (days)	5.355 to 6.858
Baseline (number of hospitalizations)	48307 to 53975
Frequency	0.01311 to 0.01428
Goodness of Fit	
Degrees of freedom	252
R squared	0.1178
Sum of squares	131250477878
Sy.x	22822
Number of Points	
# of X values	256
# Y values analyzed	256

*95% confidence interval

Midwest during the summer season. However, all regions are now showing record levels of COVID-19 cases and hospitalizations during the Winter season.

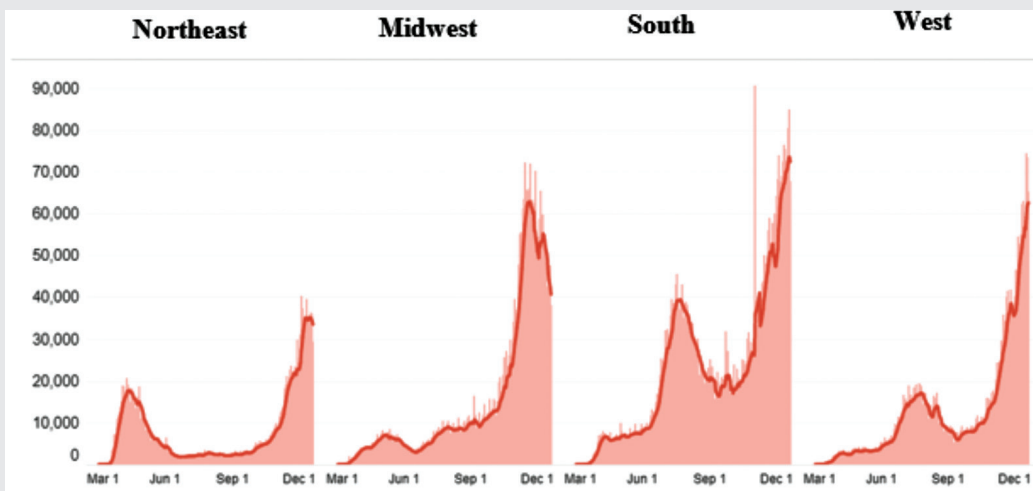


Figure 2. Regional daily COVID-19 cases from March–December 2020.

*From COVID-19 Tracking Project at the Atlantic (<https://covidtracking.com/data/national>).

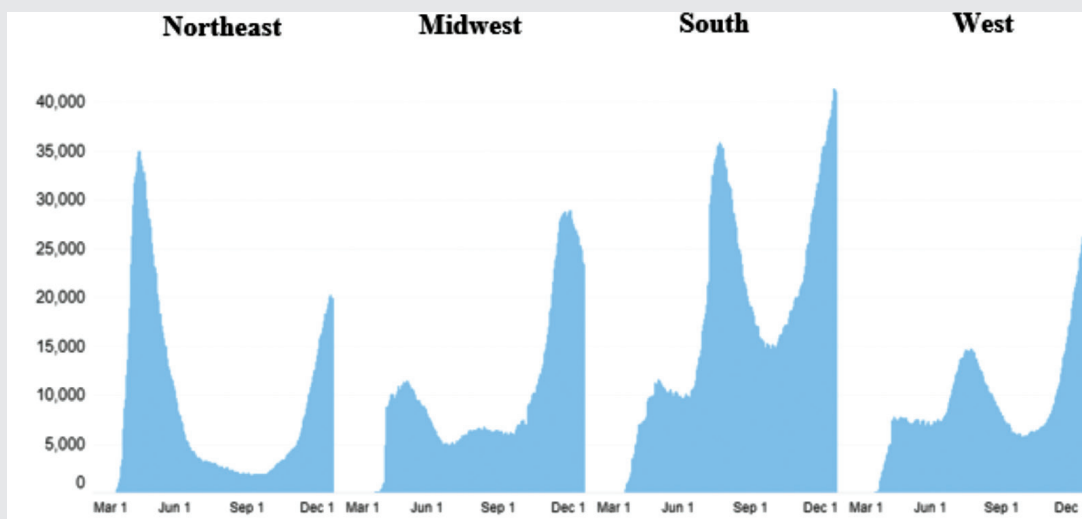


Figure 3. Daily regional COVID-19 hospitalizations from March–December 2020.
 *From COVID-19 Tracking Project at the Atlantic (<https://covidtracking.com/data/national>).

In addition, changes in hospitalizations follow increases in daily positive COVID-19 tests (Figure 5). This follows naturally with trends in hospitalization, which tend to increase within two weeks after an increase in positive COVID-19 tests. In contrast, there were higher daily COVID-19 deaths at the beginning of the pandemic than during the summer (Figure 6). From July, the daily COVID-19 deaths averaged

around 1,000 deaths per day. However, the average daily deaths in the winter season has increased to approximately 3,000 deaths per day. Furthermore, the trends in hospitalizations for the coming years remain uncertain. The recent holiday travel for Christmas and New Year celebrations may increase COVID-19 cases and hospitalizations in the month of January.

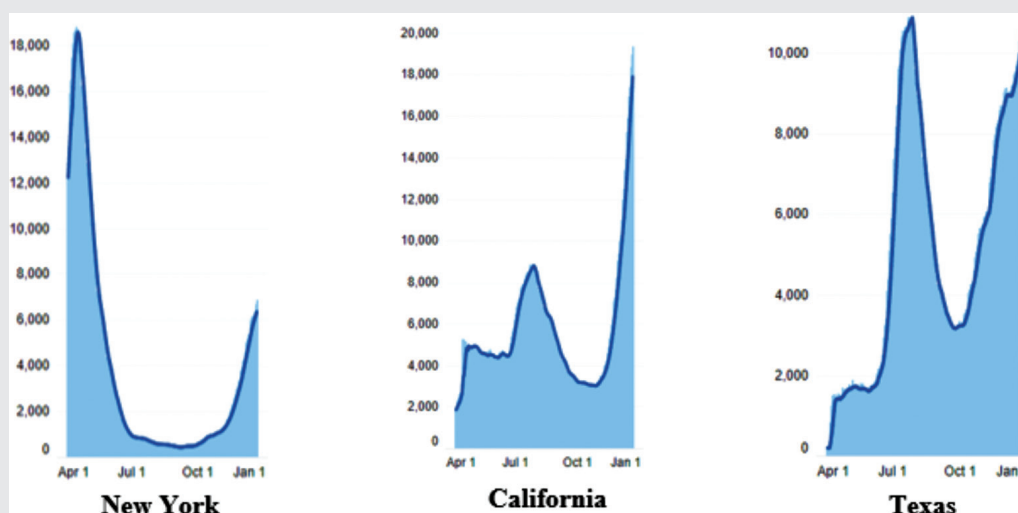


Figure 4. COVID-19 hospitalizations in New York, California, and Texas from April–December 2020.
 *From COVID-19 Tracking Project at the Atlantic (<https://covidtracking.com/data/national>).

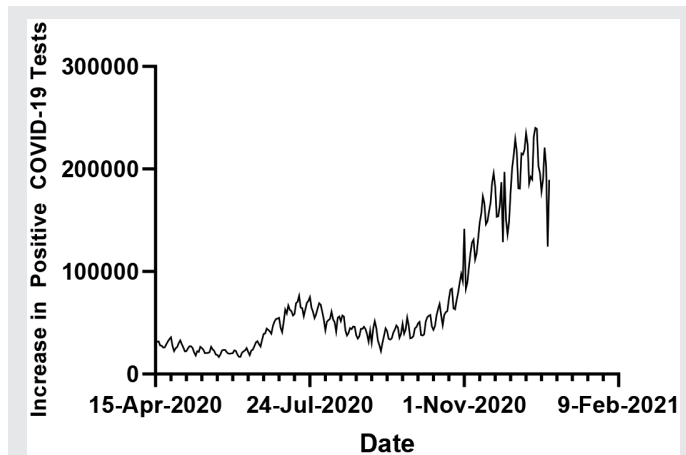


Figure 5. Daily positive COVID-19 tests from April–December 2020.

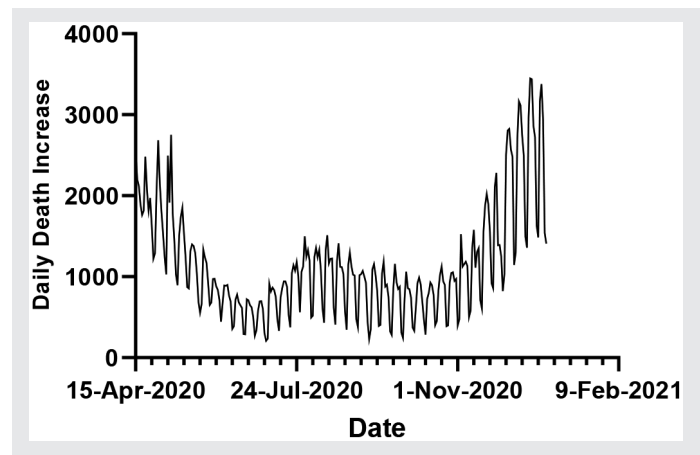


Figure 6. Daily COVID-19 deaths from April–December 2020.

DISCUSSION

Despite the regression analysis providing some insight into the trends of COVID-19 hospitalizations, the trend is likely to change with an increase in positive COVID-19 tests and hospitalizations over the fall and winter seasons. The sine wave model assumes absolute maximum and minimum values for COVID-19 hospitalizations at constant time intervals. However, the sine wave regression model was used for all COVID-19 hospitalizations in the US during the first year of the pandemic. Specifically, the sine wave analysis in this paper is the sum of all the hospitalizations across all 50 states. Furthermore, the lack of historical data makes the sine wave analysis uncertain in predicting when and how many COVID-19 hospitalizations will occur in the US. As shown in Figures 3 and 4, the sine wave regression model does not accurately model individual state hospitalizations. Therefore, the sine wave regression model works by combining state COVID-19 hospitalizations into a single variable while ignoring state-to-state variations in COVID-19 hospitalizations.

The sine wave trend in COVID-19 hospitalizations is unlikely to repeat in the future. The sine wave model assumes a fixed baseline of COVID-19 hospitalizations with a theoretical maximum and minimum values. Specifically, sine waves have upper and lower limits

that cause movement to slow, roll over in direction, and then accelerate in the new direction. Similar to an engine, this movement requires a continuous supply of energy or resources to maintain the trend. Sine waves describe processes that are inherently unstable. Furthermore, a sine wave can continue only if there is an external source of energy to replace losses dissipated during the previous cycle. It is unclear whether the COVID-19 pandemic meets either of these conditions for a sine wave.

These problems can be seen in the sine wave regression models shown in Figures 1A and B. The sine wave model predicted peaks in COVID-19 hospitalizations every 3.5 months using total COVID-19 hospitalization data from April–November 2020 (Figure 1A and Table 1A). However, the sine wave trend disappeared when the December peak in COVID-19 hospitalizations is included (Figure 1B and Table 1B). This is due to a large increase in peak total COVID-19 hospitalizations in December compared to April and July. This increase in COVID-19 hospitalizations was likely due to increased travel, more family gatherings, and increased indoor activities during the winter seasons. Similar to other viral outbreaks, such as influenza, each outbreak is due to a change in conditions, either a change in the virus, or a change in the environment favoring transmission, or both. The height of the peak is determined by the number of people

susceptible. As cases, hospitalizations, and deaths exhaust susceptible hosts, the growth slows, then rolls over, then declines. The curve does not decline to zero but persists until conditions are ripe for the next outbreak.

CONCLUSION

The continual spread of SARS-CoV-2 has driven efforts to adapt local and national responses to reduce COVID-19 cases and hospitalizations. In general, mathematical modeling of hospitalizations works best when there is an established pattern of disease transmission from multiple years of data collection; COVID-19 is a novel virus for which we have less than a year's worth of data from which to draw conclusions. Furthermore, there remains uncertainty about the trajectory of COVID-19 cases and hospitalizations in the future, particularly with the recent emergency use authorization of the Pfizer and Moderna COVID-19 vaccines. As shown in this paper, trends can appear or disappear rapidly with changes in social behavior and adherence to public health measures to prevent the spread of COVID-19.

However, the trends in COVID-19 hospitalizations provide insight into the potential risk factors influencing COVID-19 hospitalizations. The CDC data indicate that minority groups, older age, and medical comorbidities (hypertension, obesity, metabolic dysfunction, and chronic lung disease) increase the risk of hospitalization with COVID-19. Among all risk factors though, age remains the most important risk factor for COVID-19 hospitalizations and deaths. As shown in Table 2, the risk for COVID-19 hospitalizations and deaths increases proportionally compared to 18-29-year-old demographic. Beyond the age of infected COVID-19 patients, the trends in COVID-19 hospitalizations also reflect public policy with regard to re-opening the economy, school openings, social gatherings, and public fatigue with wearing masks and/or maintaining social distance. Together, these factors have a direct impact on the incidence of positive COVID-19 tests and increase the likelihood of a person's being hospitalized from COVID-19. However, it is unclear whether policies intended to decrease spread of the virus will be effective long

Table 2. Hospitalization and Death Rate Ratios Compared to 18–29 Year Olds

Age (years)	Hospitalization Rate	Death Rate
0–4	4x lower	9x lower
5–17	9x lower	16x lower
18–29	Comparison Group*	Comparison Group*
30–39	2x higher	4x higher
40–49	3x higher	10x higher
50–64	4x higher	30x higher
65–74	5x higher	90x higher
75–84	8x higher	220x higher
85+	13x higher	630x higher

¹Data source: COVID-NET (<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>, accessed 08/06/20). Numbers are unadjusted rate ratios.

²Data source: NCHS Provisional Death Counts (<https://www.cdc.gov/nchs/nvss/vsrr/COVID19/index.htm>, accessed 08/06/20). Numbers are unadjusted rate ratios.

*Units for comparisons are in mortality/hospitalization rates per 100,000 population

term or merely defer hospitalizations and deaths until the next cycle. Longer term studies are needed to determine whether additional public health measures and other factors related to hospital management may affect patients' mortality, morbidity, and quality of life during the COVID-19 pandemic.

Article citation: Kopel J, Tenner Jr TE, Gregory L, Brower GL. Modeling of COVID-19 total hospitalizations in the United States. *The Southwest Respiratory and Critical Care Chronicles* 2021;9(37):1–8

From: School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas

Submitted: 11/12/2020

Accepted: 1/13/2021

Reviewer: Gilbert Berdine MD

Conflicts of interest: none

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

REFERENCES

1. Kopel J, Goyal H, Perisetti A. Antibody tests for COVID-19. *Baylor University Medical Center Proceedings* 2020;1–10.
2. Birkmeyer JD, Barnato A, Birkmeyer N, et al. The impact Of The COVID-19 pandemic on hospital admissions in the United States. *Health Affairs* 2020;39(11):2010–2017.
3. Arcaya MC, Tucker-Seeley RD, Kim R, et al. Research on neighborhood effects on health in the United States: A systematic review of study characteristics. *Social Science & Medicine* 2016;168:16–29.
4. Mahmud N, Hubbard RA, Kaplan DE, et al. Declining cirrhosis hospitalizations in the wake of the COVID-19 Pandemic: A National Cohort Study. *Gastroenterology* 2020;159(3):1134–1136.e1133.
5. Siegler JE, Heslin ME, Thau L, et al. Falling stroke rates during COVID-19 pandemic at a comprehensive stroke center. *Journal of Stroke and Cerebrovascular Diseases* 2020;29(8):104953.
6. Solomon MD, McNulty EJ, Rana JS, et al. The Covid-19 pandemic and the incidence of acute myocardial infarction. *New England Journal of Medicine* 2020;383(7):691–693.
7. Spaccarotella CAM, De Rosa S, Indolfi C. The effects of COVID-19 on general cardiology in Italy. *European Heart Journal* 2020.
8. Ray EL, Wattanachit N, Niemi J, et al. Ensemble forecasts of coronavirus disease 2019 (COVID-19) in the U.S. *medRxiv* 2020:2020.2008.2019.20177493.