



Under-reporting of COVID-19 Hospitalizations and Its Impact on Estimating Growth Rates

Dexiang Gao^{1,2}
Stanley Xu^{2,3*}

¹Department of Pediatrics, School of Medicine, University of Colorado Anschutz medical Campus, Colorado.

²Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz medical Campus, Colorado, USA

³Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, California, USA

Abstract

Daily COVID-19 related hospitalizations have been used in monitoring the trends of the COVID-19 pandemic. However, the daily hospitalizations can be under-reported due to lags in reporting. The impact of this under-reporting on estimating the growth rate of hospitalizations has not been evaluated. Our aims of this study are 1) to investigate the under-reporting of hospitalizations in Colorado State; 2) to examine the impact of under-reporting on estimating growth rates of hospitalizations; 3) to propose a method to adjust for the under-reporting when estimating growth rates. We found that under-reporting of daily hospitalizations was severe in the early days of the pandemic, with the percentage of under-reporting as high as 59%. We also demonstrated that the under-reporting of the daily hospitalizations had the greatest impact on estimating the most recent growth rates. For the period between April 9th and April 15th, the growth rate was -0.079 (95% CIs=(-0.118, -0.041)) on report date April 15th, 2020 and it became -0.008 (95% CIs=(-0.038, 0.054)) on report date April 29th, 2020. We recommend either removal of recent data or consideration of adjusting for the under-reporting of the daily hospitalizations in estimating growth rates.

Introduction

As of June 7th, 2020, COVID-19 has infected more than 4 million people globally and killed more than 400,000. In the United States of America (USA) over 1.9 million Americans tested positive and over tens of thousands lost their lives to the disease [1]. Since the middle of March, many state governments have issued social distancing guidelines including school and non-essential business closures [2]. These mitigation measures helped flatten the curve of the pandemic, evidenced by lower daily COVID-19 hospitalizations and lower daily deaths. New York, for instance, was able to make enough progress to begin reopening their businesses on June 8th, 2020. However, some states that reopened (e.g., Texas and Arizona) saw surges in COVID-19 cases and hospitalizations two weeks after Memorial Day [3]. The numbers coming out of these states have not been the only setback. Large gatherings of demonstrators demanding social justice for African Americans across the USA have also posed a risk for increased COVID-19 transmission [4]. Given all these threats to progress on beating the epidemic, monitoring the trends of the pandemic is crucial.

The growth rate of COVID-19 cases has been used to monitor trends of the pandemic [5,6] and to evaluate the impact of social distancing guidelines in USA [7]. However, COVID-19 testing has been challenged in the USA due to the shortages in testing components. Thus, COVID-19 cases were not ideal for estimating growth rates [6]. As an alternative reliable metric, the number of COVID-19 hospitalizations was considered. Monitoring the growth rates of the hospitalizations became important for health systems to prepare and allocate limited resources [8-10].

Although some states' Department of Public Health websites report COVID-19 hospitalizations, the recent hospitalizations were often underrepresented due to lags in reporting. The objectives of this study are 1) to evaluate the under-reporting of hospitalizations; 2) to assess the impact of the under-reporting on estimating the growth rate of hospitalizations; 3) to measure the under-reporting of hospitalizations and adjust for the under-reporting of hospitalizations in estimating the growth rate.

Methods

Population and data resource

The Colorado Department of Public Health and Evaluation (CDPHE) has been collecting the cumulative numbers for COVID-19 cases, hospitalizations, and deaths on a daily basis. As of June 6th, 2020, among 5.8 million Coloradans, 210,285 people were tested; among those tested, 27,848 were COVID-19 positive, 4,480 were hospitalized, and 1,274 were dead

Article Information

DOI: 10.31021/jer.20203124
Article Type: Research Article
Journal Type: Open Access
Volume: 3 **Issue:** 2
Manuscript ID: JER-3-124
Publisher: Boffin Access Limited
Received Date: 29 June 2020
Accepted Date: 06 July 2020
Published Date: 10 July 2020

*Corresponding author:

Stanley Xu

Department of Research and Evaluation
Kaiser Permanente Southern California
100 S Los Robles Ave, Pasadena
CA 91101, USA
E-mail: Stan.Xu@kp.org

Citation: Gao D, Xu S. Under-reporting of COVID-19 Hospitalizations and Its Impact on Estimating Growth Rates. J Emerg Rare Dis. 2020 June;3(2):124.

Copyright: © 2020 Gao D, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 international License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

due to COVID-19 [11]. The CDPHE made these data available to public through a portal, which has been updated daily [12]. We considered the cumulative number of hospitalizations reported from the CDPHE as the actual numbers for those dates in Colorado.

However, due to lags in reporting from the CDPHE's partners, a cumulative number of hospitalizations for a given date was usually reported two days later. Then, it would get repeated updates over a period of time. The daily number of hospitalizations was derived by subtracting the cumulative number in the previous day from the number at the current day. We used "report date" to denote the time that the number of hospitalizations for a specific date was updated. For example, the daily numbers of hospitalizations for April 1st, 2020 were 60, 63, 70, and 80 on report dates April 3rd, 4th, 5th, and 6th, 2020, respectively. In this paper, we analyzed the numbers of hospitalizations from March 9th to June 4th, 2020.

To assess the extent of under-reporting, we first identified the time it took for the daily number of hospitalizations to reach a plateau. We then calculated the percentage of under-reporting during "the report dates" before reaching the plateau. Finally, we proposed an approach to estimate recent growth rates while adjusting for the under-reporting.

A nonlinear model to identify the time for the daily number of hospitalizations to reach a plateau

While analyzing COVID-19 hospitalizations for estimating a growth rate and doubling time [10], we observed that the cumulative number of hospitalizations was updated daily since April 1st by the CDPHE. In general, the daily number of hospitalizations increased for several days after it was first reported on the data portal. We proposed a nonlinear model to identify the time that it took (in days) for the daily number of hospitalizations to reach a plateau. The time to reach a plateau is represented by the variable t_c .

Let y_{it} be the daily number of hospitalizations for day i and reported on day t where i =March 9th to June 4th, t being the days from the date when y_{it} was first reported ($t=1$). For example, on April 6th, 30 hospitalizations for April 4th was first reported (a delay of 2 days). Over successive days, the number of hospitalizations for April 4th was updated as more reporting occurred. Finally, on May 5th, 29 days after it was first reported, the number of hospitalizations for April 4th reached its plateau. We fit separate piece-wise quadratic models to identify the time to reach a plateau ($t_{(ic)}$) for each daily number of hospitalizations y_i [13-16],

$$E(y_{it}) = \beta_{i0} + \beta_{i1}t + \beta_{i2}t^2, \text{ if } t < t_{(ic)} \text{ and } E(y_{it}) = c, \text{ if } t \geq t_{(ic)}.$$

Note that the recent daily numbers of hospitalizations may have not reached a plateau because there were more updates to come.

The percentages of under-reporting daily hospitalizations over report time, a Beta distribution and the adjusted daily number of hospitalizations for recent days

In the days before the daily hospitalizations reached a plateau, the percentage of under-reporting over time was estimated as:

$$\pi_i(t) = \frac{y_{it_{(ic)}} - y_{it}}{y_{it_{(ic)}}}, \text{ where } t < t_{(ic)}.$$

Because of lags in reporting hospitalizations, $y_{it_{(ic)}} > y_{it}$ when $t < t_{(ic)}$; thus, $\pi_i(t)$ is between 0 and 1. For each value of t , we assumed a Beta distribution for the percentages of the under-reporting across values of i ,

$$\pi(t) \sim \text{beta}(a(t), b(t)), \quad (1)$$

where $a(t)$ and $b(t)$ were the two positive shape parameters at time t for a Beta distribution [14]. It can be shown that the adjusted daily hospitalizations before reaching a plateau is

$$y_{it_adj} = \frac{y_{it}}{1 - \pi(t)}, \text{ where } t < t_{(ic)}.$$

Growth rate of the daily hospitalizations

Previously, a growth model was proposed for estimating growth rates and doubling times of the daily hospitalizations [10]. Because the growth rates and doubling times changed over time, a rolling growth curve approach (RGCA) was used with a period of 7 days. The RGCA approach excluded the recent 1-2 days' hospitalization data due to the under-reporting of hospitalizations [10]. However, that adjustment was not enough. The hospitalization data older than two days were still under-reported. Here, we propose a modified growth curve model to adjust for the under-reporting.

First without considering the lags in reporting, suppose we seek to estimate the growth rate over n days from day i (start day) to day $(i+n-1)$. Let $y_{(i+j-1)}$ denote the daily hospitalizations at day $(i+j-1)$, $1 \leq j \leq n$. Based on a growth model, we have [10]:

$$y_{(i+j-1)} = y_i(1+r)^{j-1} \quad (2)$$

where r is the growth rate during day i to day $(i+n-1)$.

Considering lags in reporting, we propose a modification to model (2):

$$\frac{y_{(i+j-1)t}}{1 - \pi(t)}(1 - I_{(i+j-1)t}) + y_{(i+j-1)t}I_{(i+j-1)t} = \left[\frac{y_{it}}{1 - \pi(t)}(1 - I_{it}) + y_{it}I_{it} \right](1 + r_a)^{j-1} \quad (3)$$

where r_a is the growth rate after adjusting for the under-reporting of hospitalizations; $I_{(i+j-1)t}$ is a binary indicator equal to 1 after $y_{(i+j-1)t}$ has reached a plateau and equal to zero if $y_{(i+j-1)t}$ has not reached a plateau; I_{it} is a different binary indicator equal to 1 when y_{it} has reached a plateau and equal to zero if y_{it} has not reached a plateau.

To estimate r_a and obtain its 95% confidence intervals (CIs), we employed the Monte Carlo (MC) simulation approach. We simulated 1000 values of $\pi(t)$ based on a Beta distribution using the shape parameters estimated in equation (1). This provided 1000 values of r_a . The average of these 1000 values of r_a was considered as the point estimate of r_a . The 2.5th and the 97.5th percentiles were considered as r_a 's 95% CIs.

Although not reported in this study, the doubling time D_a or half time H_a can be estimated by replacing r_a in equation (3) with

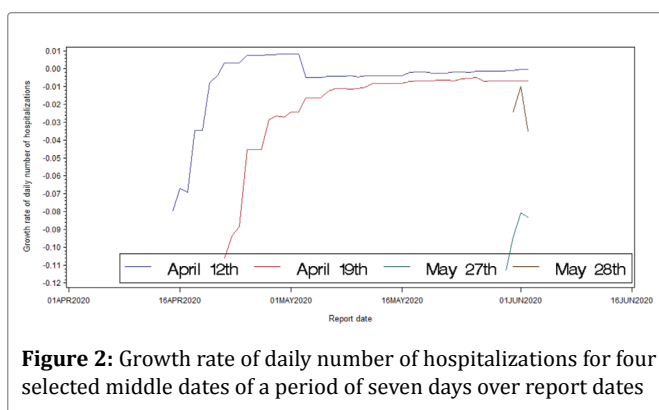
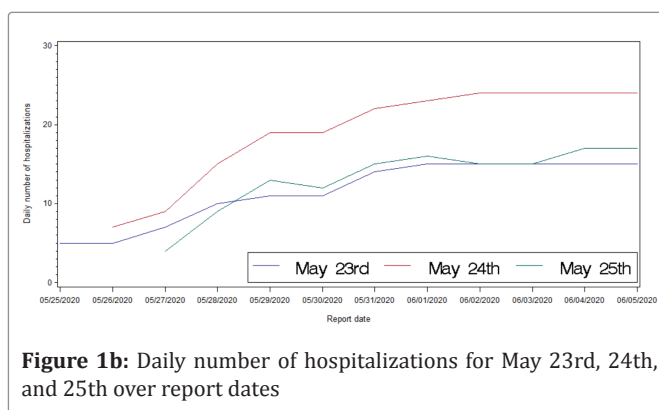
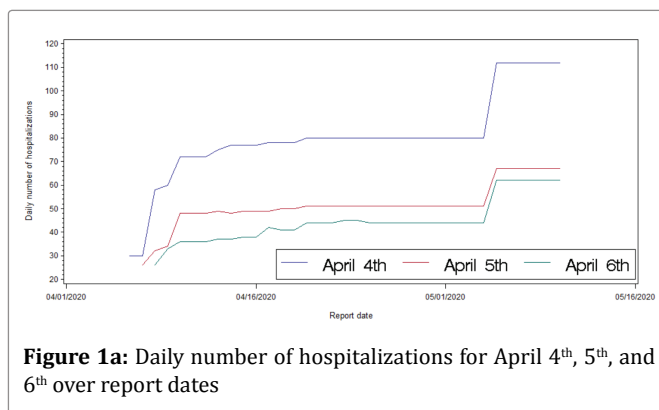
$$r_a = e^{\left(\frac{\ln(2)}{D_a - 1}\right)} - 1 \text{ when } r_a > 1 \text{ or } r_a = e^{\left(\frac{\ln(0.5)}{H_a - 1}\right)} - 1 \text{ when } r_a < 1.$$

Results

The daily number of hospitalizations before March 21st remained the same since they were first reported on the CDPHE portal on April 1st. As a result, the data before March 21st were not included in identifying the time to plateau, and not included in calculating the percentages of the under-reporting ($\pi(t)$), but were included in estimating growth rates.

Figure 1a showed the daily numbers of hospitalizations for April 4th, 5th and 6th (i) over report dates. Figure 1b showed the daily numbers of hospitalizations for May 23rd, 24th, and 25th. For the daily numbers of hospitalizations before May 5th, there were significant increases on May 5th, indicating a possible addition of partner(s) with CDPHE (Figure 1a). Both figures 1a and 1b showed a significant under-reporting of hospitalizations in the early days of reporting.

The time for the daily hospitalizations to reach a plateau was shown in groups by weeks where i date fell in (Table 1). In the early stages of the pandemic, it took a longer time for the daily hospitalizations to reach a plateau. For example, during the week of March 21-March 27, it took 37.5 days on average for the daily hospitalizations to reach a plateau; after May 2nd, it took fewer than 10 days on average for the daily hospitalizations to reach a plateau, indicating an improved reporting of daily hospitalizations in Colorado.



The percentages of under-reporting of the daily hospitalizations ($\pi(t)$) due to lags in reporting over 15 days are presented in Table 2. When daily hospitalizations were first reported on the CDPHE portal, on average they were under-reported by 59%. The percentages of under-reporting of daily hospitalizations decreased over report dates. However, even 10 days after the first reporting on the portal, the percentages of under-reporting of daily hospitalizations were still about 10%. This is an indication of the importance of adjusting for the under-reporting of recent hospitalizations for estimating recent growth rates. The two shape parameters of a Beta distribution for the percentages of under-reporting are also presented in Table 2.

In estimating the growth rates, we used the RGCA approach with a period of 7 days as described before [10]. For the daily hospitalizations in Colorado, we estimated the growth rates for the following periods during March 9th – June 4th: March 9th - 15th, 16th - 22th, 15th - 21st, ..., and May 29th - June 4th. First, we estimated the growth rates of the daily hospitalizations without adjusting for the under-reporting. Figure 2 shows the growth rates of four selected middle dates of a period of seven days over report dates. It took over two weeks for the growth rate of a seven-day period with a middle date of April 12th to stabilize: the growth rate was -0.079 (95% CIs=(-0.118, -0.041)) on report date April 15th 2020 and increased to -0.008

weeks for daily hospitalizations	Time to plateau (t_c) in days	
	Mean	Standard deviation
March 21 st - March 27 th *	37.5	5.0
March 28 th - April 3 rd	36.3	7.6
April 4 th - April 10 th	30.3	13.1
April 11 th - April 17 th	18.6	4.4
April 18 th - April 24 th	20.0	4.0
April 25 th - May 1 st	19.0	6.7
May 2 nd - May 8 th	7.9	1.1
May 9 th - May 15 th	9.7	4.1
May 16 th - May 22 nd	7.7	2.3
May 23 rd - May 29 th	5.5	1.8

* i fell between March 21st and March 27th

Table 1: Time to a plateau (t_c) by weeks for the daily hospitalizations in Colorado State

(95% CIs=(-0.038, 0.054)) on report date April 29th. The growth rates for the middle dates of May 27th and May 28th had not stabilized as of June 4th.

We then estimated the growth rates of the daily hospitalizations, adjusting for the under-reporting as proposed in METHODS. We only present the growth rates after 5/27/2020. The results were compared to those using the daily hospitalizations reported on June 6th that had hospitalization data up to June 4th. The growth rates before 5/27/2020 were the same with and without adjusting for the under-reporting because the daily hospitalizations reached a plateau and the same data were used for both approaches. The difference in growth rates was greatest for the most recent period May 29th - June 4th, 2020, the unadjusted growth rate $r = -0.14$ (95% CIs=(-0.25, -0.04)), and the adjusted growth rate $r_a = 0.10$ (95% CIs=(0.00, 0.27)) (Table 3).

Discussion

During the early days of the COVID-19 pandemic, many forecasting models painted a horrific picture for the US including Colorado [15,16]. Although those predictions did not come true, these daunting death tolls pushed local and federal governments to form and implement social distancing guidelines which in turn helped flatten the pandemic curves in states such as Colorado, New York, and other states. When COVID-19 cases increased in the US, some researchers started to use available data to inform models so that their forecasts were more relevant to the US population. However, these numbers can be under-reported and this under-reporting may have contributed to the inaccuracy of forecasting. To our knowledge, this is the first study that has examined the under-reporting of COVID-19 related hospitalizations and its impact on estimating growth rates.

There are several limitations in this study. First, we used only Colorado's hospitalization data and the results may not be generalizable to other regions. However, we expect similar under-reporting issues in other states. Second, the percentages of the under-reporting of daily hospitalizations ($\pi(t)$) were estimated using data including those from the early pandemic when the under-reporting was severe. They may not represent the under-reporting of recent daily hospitalizations. As a result, the recent growth rates may be overcorrected. A solution is to use the more recent data (e.g., data during May 2020) to estimate the percentages of the under-reporting ($\pi(t)$).

We conclude that it is crucial to adjust for the under-reporting of hospitalizations to monitor the trend of the pandemic in real-time. If it takes a few days for the daily hospitalization number to reach a plateau, we recommend removal of recent data in estimating growth rates; if it takes a longer time for number to reach their final values, we recommend estimating the percentages of under-reporting of daily

the days from the date when y_{it} was first reported (t)	percentages of under-reporting ($\pi(t)$)			Shape parameters for a Beta distribution	
	Mean	Standard deviation	Median	a	b
1	0.59	0.14	0.61	6.28	4.37
2	0.50	0.15	0.49	4.92	4.74
3	0.41	0.15	0.40	4.08	5.80
4	0.33	0.12	0.33	4.19	8.39
5	0.26	0.09	0.27	5.04	14.16
6	0.20	0.09	0.21	3.42	13.29
7	0.16	0.08	0.16	2.82	13.95
8	0.14	0.08	0.12	2.65	15.16
9	0.13	0.08	0.12	2.35	14.61
10	0.13	0.07	0.12	2.91	17.97
11	0.13	0.07	0.12	2.85	18.46
12	0.12	0.06	0.12	2.62	18.93
13	0.11	0.06	0.12	2.71	20.62
14	0.11	0.06	0.11	2.78	22.07
15	0.10	0.06	0.10	2.80	22.82

Table 2: The percentages of under-reporting of the daily hospitalizations ($\pi(t)$) due to lags in reporting over time from the date when y_{it} was first reported (t) and the shape parameters of a Beta distribution

The middle day of a period of seven days	Unadjusted		adjusted	
	Growth rate (r)	95% CIs	Growth rate (r_a)	95% CIs
5/27/2020	-0.07	(-0.14, 0.00)	-0.07	(-0.07, -0.05)
5/28/2020	-0.03	(-0.12, 0.07)	0.00	(-0.02, 0.04)
5/29/2020	-0.10	(-0.17, -0.03)	-0.03	(-0.07, 0.05)
5/30/2020	-0.16	(-0.23, -0.09)	-0.01	(-0.08, 0.09)
5/31/2020	-0.15	(-0.23, -0.06)	0.03	(-0.05, 0.16)
6/1/2020	-0.14	(-0.25, -0.04)	0.10	(0.00, 0.27)

Table 3: Growth rates of the daily hospitalizations due to COVID-19 in Colorado State with and without adjusting for the under-reporting of hospitalizations

hospitalizations and application of the newly proposed approach; if the percentages of under-reporting of hospitalizations changes over time, we recommend use of recent data to estimate the percentages of under-reporting of daily hospitalizations and application of the newly proposed approach.

Rererences

1. CDC. accessed on 6/7/2020
2. <https://web.csg.org/covid19/executive-orders/> accessed on 6/7/2020
3. <https://www.cnn.com/2020/06/10/health/us-coronavirus-wednesday/index.html> accessed on 6/12/2020
4. <https://www.washingtonpost.com/health/2020/06/04/cdc-director-says-protesters-should-consider-getting-tested-covid-19/> Accessed on 6/12/2020.
5. Leclerc Q, Nightingale ES, Abbott S, CMMID nCov working group & Jombart T. Analysis of temporal trends in potential COVID-19 cases reported through NHS Pathways England. <https://cmmid.github.io/topics/covid19/nhs-pathways.html> accessed on 6/12/2020
6. Omori R, Mizumoto K, and Chowell G. Changes in testing rates could mask the novel coronavirus disease (COVID-19) growth rate. *Int J Infect Dis*. 2020 May;94:116–118.
7. Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate.
8. Karaca-Mandic P, Georgiou A, Sen S. Calling all states to report standardized information on COVID-19 hospitalizations. Health Affairs blog. April 7, 2020. Accessed June 11, 2020.
9. Sen S, Karaca-Mandic P, Georgiou A. Association of Stay-at-Home Orders With COVID-19 Hospitalizations in 4 States. *JAMA*. 2020 May;323(24):2522-2524.
10. Xu S, Clarke C, Shetterly S, Narwaney K. Estimating the Growth Rate and Doubling time for Short-Term Prediction and Monitoring Trend During the COVID-19 Pandemic with a SAS Macro. *J Emerg Rare Dis*. 2020 May;3(1):121.
11. Colorado COVID-19 Updates. Data. accessed on June7th 2020
12. https://drive.google.com/drive/folders/1bBAC7H-pdEDgPxRuU_eR36ghzc0HWNf1
13. <https://support.sas.com/documentation/onlinedoc/stat/131/nlin.pdf> accessed on may 20th, 2020
14. Johnson N L; Kotz S; Balakrishnan N. (1995). "Chapter 21:Beta Distributions". *Continuous Univariate Distributions Vol. 2* (2nd ed.). Wiley. ISBN 978-0-471-58494-58500.
15. IHME COVID-19 health service utilization forecasting team. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by US state in the next 4 months. *MedRxiv*. 2020 Mar.
16. Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imp Coll COVID-19 Response Team*. 2020 Mar:20.