


# Ensuring trust in COVID-19 data

## A retrospective cohort study

Daniel Antwi-Amoabeng, MD, MSc<sup>a</sup>, Bryce D. Beutler, MD<sup>a,\*</sup> , Gurpreet Chahal, MD<sup>a</sup>, Sumaiya Mahboob, MD<sup>a</sup>, Nageshwara Gullapalli, MD, MPH<sup>a</sup>, Rudy Tedja, DO<sup>a,b,c</sup>, Farah Madhani-Lovely, MD<sup>a,b</sup>, Chris Rowan, MD<sup>a,d</sup>

### Abstract

There are no standardized methods for collecting and reporting coronavirus disease-2019 (COVID-19) data. We aimed to compare the proportion of patients admitted for COVID-19-related symptoms and those admitted for other reasons who incidentally tested positive for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

Retrospective cohort study

Data were sampled twice weekly between March 26 and June 6, 2020 from a “COVID-19 dashboard,” a system-wide administrative database that includes the number of hospitalized patients with a positive SARS-CoV-2 polymerase chain reaction test. Patient charts were subsequently reviewed and the principal reason for hospitalization abstracted.

Data collected during a statewide lockdown revealed that 92 hospitalized patients had positive SARS-CoV-2 test results. Among these individuals, 4.3% were hospitalized for reasons other than COVID-19-related symptoms but were incidentally found to be SARS-CoV-2-positive. After the lockdown was suspended, the total inpatient census of SARS-CoV-2-positive patients increased to 128, 20.3% of whom were hospitalized for non-COVID-19-related complaints.

In the absence of a statewide lockdown, there was a significant increase in the proportion of patients admitted for non-COVID-19-related complaints who were incidentally found to be SARS-CoV-2-positive. In order to ensure data integrity, coding should distinguish between patients with COVID-19-related symptoms and asymptomatic patients carrying the SARS-CoV-2 virus.

**Abbreviations:** COVID-19 = coronavirus disease-2019, ICD-10 = International Classification of Diseases, Tenth Edition, SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2.

**Keywords:** coronavirus, coronavirus disease, coronavirus disease-2019, epidemiology, health systems, severe acute respiratory syndrome coronavirus 2

## 1. Introduction

Global prediction models indicate that a fourth wave of coronavirus disease-2019 (COVID-19), driven by the emerging delta variant, may be approaching. Accurate epidemiologic data is therefore of paramount importance in modeling resource

utilization and health care systems planning. At present, there are no standardized methods for collecting and reporting COVID-19 data. In light of growing distrust in central governments to effectively manage this global crisis, data integrity has become all the more important to increase confidence in public health interventions.<sup>[1]</sup> Indeed, even clinicians and public health experts have begun to question the quality of COVID-19 data that have been published in the medical literature<sup>[2]</sup>; recent study retractions have done little to alleviate these concerns.<sup>[3]</sup>

A central challenge of COVID-19 analytics is related to the high proportion of asymptomatic infections; morbidity and mortality data is inextricably linked to the number of incidental positive severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) tests among patients hospitalized for conditions unrelated to COVID-19. We aimed to investigate the distribution of patients who were admitted for COVID-19 symptoms versus those admitted for unrelated reasons who were incidentally found to have COVID-19. In addition, we compared data obtained during a statewide lockdown with that obtained in the postlockdown period to determine the effect of community spread on the rate of incidental positive tests.

## 2. Methods

Between March 26 and June 6, 2020, we sampled data from our “COVID-19 Dashboard,” a system-wide administrative database that includes the number of hospitalized patients with a positive SARS-CoV-2 polymerase chain reaction test. We defined

Editor: Maya Saranathan.

The authors have no funding and conflicts of interest to disclose.

This study was exempt from approval by the Institutional Review Board, as it was conducted as a quality improvement project.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

<sup>a</sup> Department of Internal Medicine, University of Nevada Reno School of Medicine, Reno, NV, <sup>b</sup> Department of Pulmonary and Critical Care Medicine, Renown Health, Reno, NV, <sup>c</sup> Department of Infection Prevention, Renown Health, Reno, NV, <sup>d</sup> Institute for Heart and Vascular Health, Renown Regional Medical Center, Reno, NV.

\* Correspondence: Bryce D. Beutler, 1155 Mill Street, W-11, Reno, NV 89052 (e-mail: brycebeutler@hotmail.com).

Copyright © 2021 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Antwi-Amoabeng D, Beutler BD, Chahal G, Mahboob S, Gullapalli N, Tedja R, Madhani-Lovely F, Rowan C. Ensuring trust in COVID-19 data: a retrospective cohort study. *Medicine* 2021;100:35(e26972).

Received: 29 October 2020 / Received in final form: 6 July 2021 / Accepted: 1 August 2021

<http://dx.doi.org/10.1097/MD.00000000000026972>

“incidental positive cases” as patients hospitalized for non-COVID-related conditions who tested positive for SARS-CoV-2 but were represented as hospitalized COVID-19 patients in the database. Using the Mann–Whitney *U* test, we compared the mean daily frequency of incidental positive cases during and after a statewide lockdown, which was imposed between March 26 and April 30 and during which time all nonessential businesses were required to close and nonemergent surgical procedures were postponed. The difference in the proportion of incidental positive cases during and after the lockdown was assessed using a 2-sample Chi square test of proportions at 95% confidence. Stata version 16 (StataCorp., College Station, TX) was used for the statistical analysis.

The study was exempt from approval by the Institutional Review Board, as it was conducted as a quality improvement project.

### 3. Results

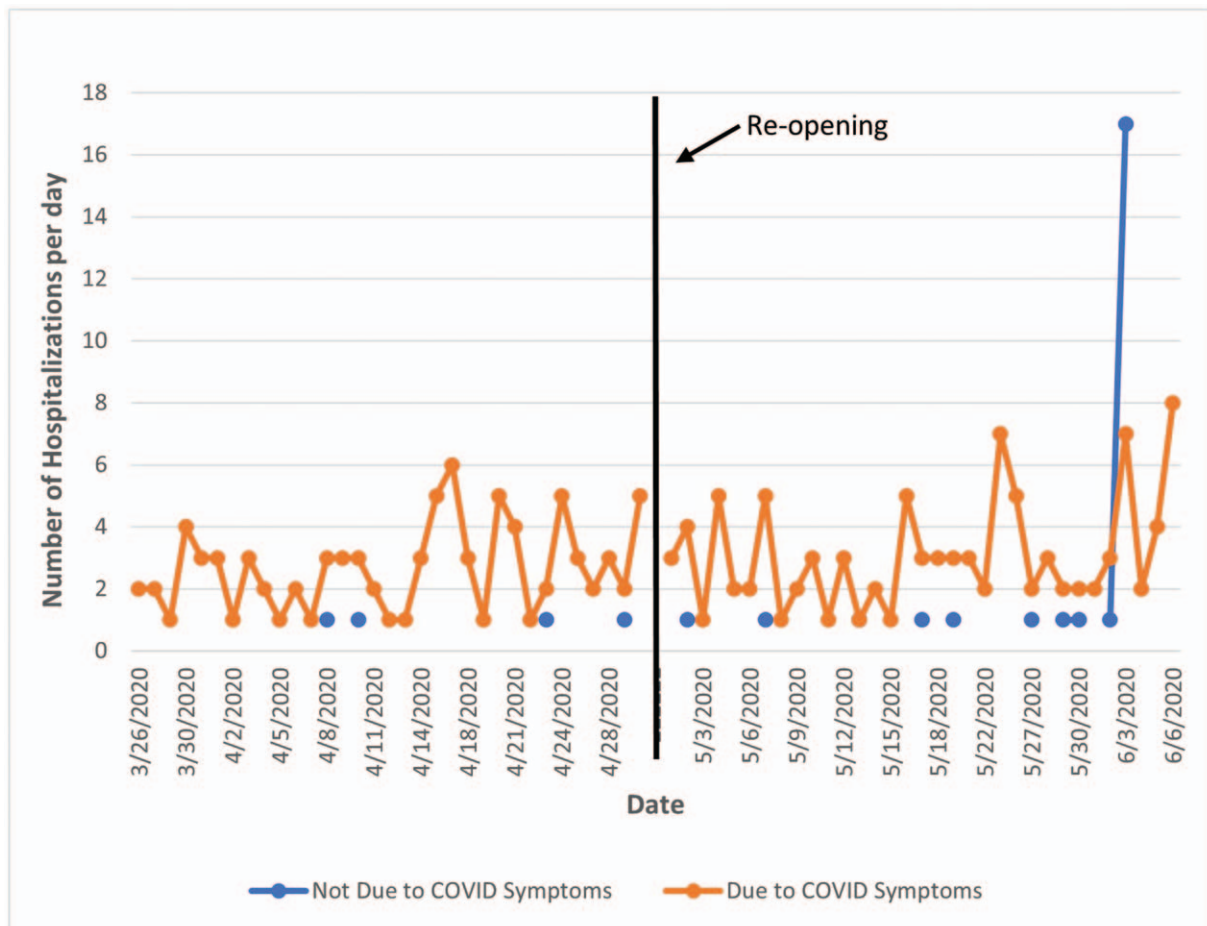
This study included 220 hospitalizations between March 26 and June 6, 2020. All patients were coded as being hospitalized for COVID-19. The total rate of incidental SARS-CoV-2 positivity in the cohort was 13.6%.

Data collected between March 26 and April 30 (33 days of lockdown) revealed that 92 hospitalized patients had positive SARS-CoV-2 test results. Among these individuals, 4.3% were hospitalized for reasons other than COVID-19 but were incidentally found to be SARS-CoV-2-positive. Between May 1 and June 6 (33 days of re-opening), the total inpatient census of SARS-CoV-2-positive patients increased to 128, 20.3% of whom were hospitalized for non-COVID-19-related complaints.

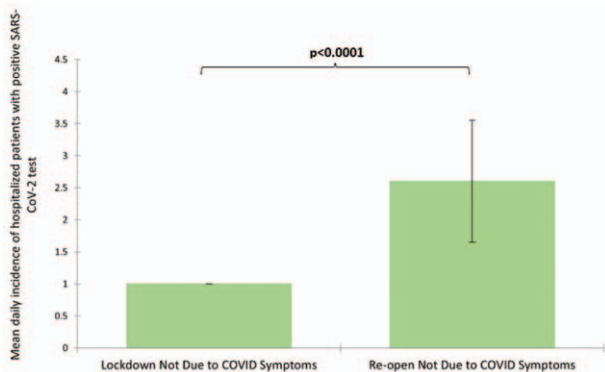
The daily trend in SARS-CoV-2-positive patients remained stable in the lockdown period but increased approximately 2 weeks after the re-opening (Fig. 1). The mean number of daily non-COVID-19-related admissions was significantly higher after the re-opening period as compared to the lockdown period (2.6 versus 1, respectively; Fig. 2). In addition, there was a statistically significant ( $P = .001$ ) increase in the proportion of incidental positive SARS-CoV-2 tests between the lockdown and immediate postlockdown periods (4.3% and 20.3%, respectively). The relationship between timeline and hospitalization type is further illustrated in Table 1.

### 4. Discussion

In June 2020, the National Institutes of Health announced the creation of the National COVID Cohort Collaborative, which



**Figure 1.** Trends in the daily number of hospitalizations of SARS-CoV-2 positive patients due to COVID symptoms versus not due to COVID symptoms. Vertical line (arrow) denotes the transition from the lockdown period (March 24<sup>th</sup> to April 30<sup>th</sup>) and the re-opening period (May 1<sup>st</sup> to June 6<sup>th</sup>). COVID = coronavirus disease, SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2.



**Figure 2.** Comparison of the mean number of daily hospitalizations (error bars = confidence intervals) of patients who did not have COVID-19 symptoms during the lockdown period (March 24<sup>th</sup> to April 30<sup>th</sup>) and the re-opening (May 1<sup>st</sup> to June 6<sup>th</sup>). There were significantly more incidental SARS-CoV-2 hospitalization during the re-opening period (2.6 versus 1.0). COVID-19 = coronavirus disease-2019, SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2.

aims to collect and study medical record data on persons diagnosed with COVID-19.<sup>[4]</sup> Although the National COVID Cohort Collaborative is a laudable undertaking, the quality of the collected data will determine the utility of the database as well as the external validity of findings from studies using the database. We have concerns about the variability in data collection across institutions and are wary of coding practices that do not distinguish asymptomatic patients from those hospitalized with COVID-related symptoms. Here, we describe our experience from our hospital in northern Nevada.

We investigated incidence patterns of COVID-19 among patients in our 808-bed tertiary referral center to assess the distribution of patients who were admitted *for* COVID-19 versus those *with* an incidental finding of COVID-19. We hypothesized that universal testing of patients presenting to the emergency department would result in a higher number of positive cases among hospitalized patients. We further hypothesized that the proportion of patients admitted with a diagnosis of COVID-19 – with or without COVID-19-related symptoms – would increase in tandem with the prevalence of the disease in the community.

The increase in the number of hospitalized patients with positive SARS-CoV-2 tests results can likely be attributed to rising community prevalence in the setting of increased admissions for non-COVID-19 complaints during the postlockdown period; data show that emergency department visits were markedly decreased during state and national lockdowns, likely

due to patient concerns for COVID-19 exposure.<sup>[5]</sup> During the lockdown and immediate postlockdown periods, our hospital instituted a policy of SARS-CoV-2 screening for all patients who were to be hospitalized. A consequence of this policy was an increase in the number of positive cases among hospitalized patients. Examples include a teenager with a traumatic femur fracture, a primigravid patient in labor, and a middle-aged man with appendicitis, all of whom were incidentally found to be SARS-CoV-2-positive but were erroneously represented as hospitalized COVID-19 patients in the hospital administrative database.

Two new provisional International Classification of Diseases, Tenth Edition (ICD-10) codes are used for establishing a diagnosis of COVID-19: U07.1, which denotes positive identification of the SARS-CoV-2 virus and U07.2, which indicates no laboratory confirmation of the virus in the setting of a clinical diagnosis of COVID-19. These codes are utilized for both clinical and epidemiological coding of the disease and will be used to monitor incidence, prevalence, hospitalizations, and in-hospital outcomes.<sup>[6]</sup> Since up to 30% of patients who test positive for SARS-CoV-2 may be asymptomatic, hospital data standardization is vital to clearly distinguish incidental from symptomatic cases. Indeed, there is a need to define these codes to indicate which patients suffered significant clinical effects of COVID-19, as the current recommendation is to use additional diagnoses to identify disease manifestations.<sup>[7]</sup> It will also be prudent to define ICD-10 codes for organ-specific COVID-19 dysfunction in affected patients.

Notably, data suggest that a significant proportion of COVID-19 cases go unreported<sup>[8,9]</sup>; this affects calculation of the R<sub>0</sub> as well as the case fatality rate. Estimates of the true disease burden rely on correlation of symptomatic, asymptomatic, and unreported cases, and thus it is essential to define ICD-10 codes appropriately to distinguish between incidental and nonincidental cases.

This report draws attention to inherent flaws in the current practice of collecting administrative data on COVID-19 hospitalizations. We believe the practices described herein are representative of the general United States hospital setting. However, we are limited by our inability to include data from other hospitals in our analysis. Further, our sample size is rather small and may underestimate the proportion of hospitalizations with SARS-CoV-2 infection without clinical evidence of COVID-19 disease.

### 5. Conclusion

Scientific evidence represents the foundation of all effective public health policy. Implementation of population-level interventions often require years of epidemiologic data collection. The COVID-19 pandemic is rapidly evolving as new variants emerge, and the medical community is frantically trying to stem the spread of the disease while collecting data to inform public health interventions. However, ensuring data integrity is crucial. We must establish clear definitions of disease states and associated outcomes in order to safeguard clinical and epidemiological research; we believe that this will ultimately strengthen interventions to control the disease. We need to change reporting to designate which patients are admitted *for* COVID-19 and which are admitted *with* COVID-19. Furthermore, designating the extent of organ involvement will be valuable in monitoring long-term sequelae of COVID-19. Public trust must be regained in our collective efforts to manage this pandemic.

**Table 1**  
**Cross-tabulation of COVID-19 hospitalization type with respect to the lockdown timeline. There was an association between the timeline and hospitalization type (P-value for  $\chi^2$  test of association < .001).**

Timeline	Type of COVID-19 hospitalization		
	Incidental	Symptomatic	Row total
Lockdown	4 (4.3%)	88 (95.7%)	92 (100%)
Re-opening	26 (20.3%)	102 (79.7%)	128 (100%)
Row total	30 (13.6%)	190 (86.4%)	220 (100%)

COVID-19 = coronavirus disease-2019.

## Author contributions

**Conceptualization:** Daniel Antwi-Amoabeng, Gurpreet Chahal, Nageshwara Gullapalli.

**Data curation:** Daniel Antwi-Amoabeng, Sumaiya Mahboob, Rudy Tedja.

**Funding acquisition:** Gurpreet Chahal.

**Investigation:** Daniel Antwi-Amoabeng, Bryce David Beutler, Rudy Tedja, Farah Madhani-Lovely.

**Methodology:** Rudy Tedja.

**Supervision:** Chris Rowan.

**Writing – original draft:** Bryce David Beutler.

**Writing – review & editing:** Nageshwara Gullapalli.

## References

- [1] Peretti-Watel P, Seror V, Cortaredona S, et al. A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicization. *Lancet Infect Dis* 2020;20:769–70.
- [2] Ray Kalyan. Scientists Raise Questions on Quality of COVID-19 Data Flaunted by Centre as Infections Soar. *Deccan Herald, DH News Service* 2020;19 May 2020, [www.deccanherald.com/national/scientists-raise-questions-on-quality-of-covid-19-data-flaunted-by-centre-as-infections-soar-839141.html](http://www.deccanherald.com/national/scientists-raise-questions-on-quality-of-covid-19-data-flaunted-by-centre-as-infections-soar-839141.html).
- [3] Ledford H, Van Noorden R. High-profile coronavirus retractions raise concerns about data oversight. *Nature* 2020;582:160.
- [4] Rubin R. NIH launches platform to serve as depository for COVID-19 medical data. *JAMA* 2020;324:326.
- [5] Czeisler MÉ, Marynak K, Clarke KEN, et al. Delay or avoidance of medical care because of COVID-19-related concerns – United States, June 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:1250–7.
- [6] World Health Organization . ‘Emergency Use ICD Codes for COVID-19 Disease Outbreak. 2020;World Health Organization, 22 May 2020, [www.who.int/classifications/icd/covid19/en/](http://www.who.int/classifications/icd/covid19/en/).
- [7] World Health Organization . International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)-WHO Version for 2019. 2019;World Health Organization, 30 May 2020, <https://icd.who.int/browse10/2019/en#U07.1>.
- [8] Maugeri A, Barchitta M, Battiato S, Agodi A. Estimation of unreported SARS-CoV-2 cases in Italy using a susceptible-exposed-infectious-recovered-dead model. *J Glob Health* 2020;10:21105.
- [9] Maugeri A, Barchitta M, Battiato S, Agodi A. Estimation of unreported novel coronavirus (SARS-CoV-2) infections from reported deaths: a susceptible-exposed-infectious-recovered-dead model. *J Clin Med* 2020;9:1350.