

Validation of Real-World Case Definitions for COVID-19 Diagnosis and Severe COVID-19 Illness Among Patients Infected with SARS-CoV-2: Translation of Clinical Trial Definitions to Real-World Settings

Running title: Real-world validation of severe COVID-19 illnesses

Mei Sheng Duh, MPH, ScD;¹ Catherine Nguyen, MPH;² Heather Rubino, MS, PhD;³ Christopher Herrick, MBA;⁴ Rose Chang MSPH, MS, ScD;¹ Maral DerSarkissian, PhD;² Yichuan Grace Hsieh, PhD, RN;⁴ Azeem Banatwala, BA, BS;² Louise H. Yu, MS;⁵ Gregory Belsky, BS;⁴ Marykate E. Murphy, RN;⁴ Janet Boyle-Kelly, RN;⁴ Andrew Cagan, BS;⁴ Bruce E. Stangle, PhD;¹ Pierre Y. Cremieux, PhD;¹ Francesca Kolitsopoulos, MPH;³ Shawn N. Murphy, MD, PhD⁴

¹Analysis Group, Inc., Boston, MA

²Analysis Group, Inc., Los Angeles, CA

³Pfizer, Inc., New York, NY

⁴Mass General Brigham, Boston, MA

⁵Analysis Group, Inc., Montreal, Canada

*Corresponding author:

Mei Sheng Duh, MPH, ScD
Managing Principal
Analysis Group, Inc.
111 Huntington Avenue, 14th Floor
Boston, MA 02199
Tel: 617 425 8131
Mei.Duh@analysisgroup.com

ABSTRACT

Purpose: This study assessed the performance of International Classification of Diseases 10th Revision, Clinical Modification (ICD-10-CM) coronavirus disease 2019 (COVID-19) diagnostic code U07.1 against polymerase chain reaction (PCR) test results (Objective 1), and electronic medical record (EMR)-based codified algorithm for severe COVID-19 illness based on endpoints used in the Pfizer-BioNTech COVID-19 vaccine trial against chart review (Objective 2).

Methods: This retrospective, longitudinal cohort study used EMR data from the Mass General Brigham COVID-19 Data Mart (3/1/2020–11/19/2020) for adult patients with ≥ 1 PCR test, antigen test, or code U07.1 (Objective 1) and adult patients with a positive PCR test hospitalized with COVID-19 (Objective 2).

Results: Among 354,124 patients in Objective 1, 96% had ≥ 1 PCR test (including 6% with ≥ 1 positive PCR test; 11% with ≥ 1 code U07.1). Code U07.1 had low sensitivity (54%) and positive predictive value (PPV; 63%) but high specificity (97%) against the PCR test. Among 300 patients hospitalized for COVID-19 randomly sampled for chart review in Objective 2, the EMR-based case definition for severe COVID-19 illness had high PPV ($>95\%$), showing better performance than severe/critical COVID-19 endpoints defined by the World Health Organization (PPV: 79%).

Conclusions: COVID-19 diagnosis based on ICD-10-CM code U07.1 had inadequate sensitivity and requires confirmation by PCR testing. The EMR-based case definition showed high PPV and can be used to identify cases of severe COVID-19 illness in real-world datasets. These findings highlight the importance of validating outcomes in real-world data, and can guide researchers analyzing COVID-19 data when PCR tests are not readily available.

Key words: Severe COVID-19 illness; ICD-10-CM diagnosis code U07.1; PCR test; Electronic medical record

KEY POINTS

- This study evaluated the performance of International Classification of Diseases 10th Revision, Clinical Modification (ICD-10-CM) codes and an electronic medical record (EMR)-based algorithm for identifying coronavirus disease 2019 (COVID-19) diagnosis and severe COVID-19 illness in real-world data.
- ICD-10-CM code U07.1 for COVID-19 had low sensitivity and positive predictive value (PPV) against PCR tests.
- The EMR-based algorithm for severe COVID-19 illness developed from the Pfizer–BioNTech COVID-19 vaccine trial had high PPV against chart review, and may be used to identify severe cases in real-world data.
- These results highlight the importance of validating outcomes when conducting analyses of real-world datasets.

PLAIN LANGUAGE SUMMARY

As polymerase chain reaction (PCR) tests for coronavirus disease 2019 (COVID-19) diagnosis are becoming less frequently used and there is no standard definition of severe COVID-19 illness, it is important to have a way of correctly identifying COVID-19 diagnosis or severe COVID-19 illness in real-world data (e.g., electronic medical records [EMRs]). This study examined: 1) how a diagnosis code for COVID-19 used in EMRs (i.e., U07.1) compares to PCR test results in terms of accurately identifying patients with COVID-19; and 2) whether a definition for severe COVID-19 illness developed based on the Pfizer–BioNTech COVID-19 vaccine trial and a definition used by the World Health Organization [WHO] can be used to

accurately identify patients with severe COVID-19 illness in EMRs. The results showed that code U07.1 was not very accurate in identifying patients with COVID-19. On the other hand, the developed definition for severe COVID-19 illness was more accurate than the WHO definition and was able to identify most patients with severe COVID-19 illness in real-world data.

INTRODUCTION

Several vaccines have been developed to respond to the coronavirus disease 2019 (COVID-19) pandemic following the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which was identified in the United States (US) in March 2020.¹ For a COVID-19 vaccine to be approved by the US Food and Drug Administration (FDA), it must demonstrate $\geq 50\%$ efficacy compared with placebo in preventing COVID-19 illness.^{2,3} In the clinical trial for the SARS-CoV-2 RNA vaccine BNT162b2 developed by Pfizer in collaboration with BioNTech (NCT04368728), the vaccine showed 95% efficacy against COVID-19 and 90% efficacy against severe COVID-19 illness, and was well tolerated.^{4,5} Based on these results, BNT162b2 was granted emergency use authorization (EUA) on December 11, 2020⁶ and full approval on August 23, 2021.⁷

Despite the availability of vaccines, only about two-thirds (67.5%) of the US population was fully vaccinated as of September 6, 2022.⁸ As new SARS-CoV-2 variants have emerged, unvaccinated persons have shown consistently higher rates of hospitalization and death from COVID-19 than those who are vaccinated.^{9,10} Early identification and treatment of severe cases can prevent clinical progression¹¹ but remain challenging. Reverse transcription polymerase chain reaction (PCR) tests are considered the gold standard diagnostic method¹² but are resource-intensive and therefore not ideal for point-of-care testing. Moreover, there is no consistent definition of COVID-19 illness severity. The development and validation of COVID-19 diagnostic measures and severe illness endpoints used outside of clinical trials that can accurately identify these measures in varied sources such as electronic medical records (EMRs) and claims databases are important for assessing the efficacy and safety of COVID-19 vaccines and treatments in real-world settings.

As the performance of the International Classification of Diseases 10th Revision, Clinical Modification (ICD-10-CM) diagnosis code U07.1 may vary over time and by population, the performance of the code was validated against the gold-standard PCR test (Part 1). Additionally, a case definition of severe COVID-19 illness in patients hospitalized with COVID-19 using EMR data was evaluated by medical chart review (Part 2).

METHODS

Study design

A retrospective, longitudinal cohort study was conducted to evaluate COVID-19 diagnostic measures. The index date was defined as the date of the earliest COVID-19 diagnostic measure (i.e., PCR molecular test, diagnosis code U07.1, or antigen test) (Part 1). To assess the performance of an EMR-based case definition for severe COVID-19 illness, the index date was defined as the date of hospitalization with COVID-19 (readmission within 3 days was considered a continuation of the previous hospitalization), and the baseline period was defined as the 3-month period preceding the index date (Part 2). The observation period in both parts of the study was the time from the index date to the earliest of death, loss to follow-up, or end of data availability.

Data source and extraction

The Mass General Brigham (MGB) Research Patient Data Registry COVID-19 Data Mart provided deidentified inpatient and outpatient EMR data from the MGB healthcare system for more than 6.5 million patients with over 2 billion records. The MGB system consists of 8 major hospitals affiliated with Harvard Medical School in Massachusetts (Massachusetts General Hospital, Brigham and Women's Hospital, Brigham and Women's Faulkner Hospital,

Massachusetts Eye and Ear Hospital, McLean Hospital, Newton-Wellesley Hospital, North Shore Medical Center, and Spaulding Rehabilitation Hospital).¹³ Data elements include demographics, providers, visits, diagnoses, medications, procedures, laboratory, microbiology, and work-up reports (e.g., discharge, operative, radiology). Codified EMR data from March 1, 2020 (when PCR test results were first available) to November 19, 2020 were extracted for this study.

Study population

Part 1 of the study included patients aged 18–85 years with ≥ 1 diagnostic measure of COVID-19 at the time the measure was recorded. All patients were included in the calculation of frequency of diagnostic measures, but only those with PCR test results and diagnosis code U07.1 were included in the validation analysis. Part 2 included patients aged 18–85 years with a positive PCR test and hospitalization associated with COVID-19 (defined as ≥ 1 positive PCR test in the 14 days prior to hospital admission or ≥ 1 positive PCR test in the 2 days following hospital admission).

Severe and critical COVID-19 illness endpoint definitions

Severe COVID-19 illness endpoints were defined using an algorithm derived from the Pfizer-BioNTech COVID-19 vaccine trial⁵ that included diagnosis codes, procedure codes, laboratory test results, and medication records. Endpoints used by the WHO¹⁴ to classify COVID-19 illness as severe or critical were also similarly identified. The EMR-based case definition was validated through medical chart review in a randomly selected sample of 300 patients with and 100 patients without severe COVID-19 illness in MGB's codified data according to the Pfizer-BioNTech COVID-19 vaccine trial definition. Severe or critical COVID-19 illness based on the

WHO definition was also validated through medical chart review for the subsample of patients who met both the vaccine trial and WHO definitions of severe COVID-19 illness.

Statistical analysis

Data were analyzed using SQL Server Management Studio 18 (Microsoft, Redmond, WA, USA) and SAS Enterprise Guide v7.1 (SAS Institute, Cary, NC, USA).

For Part 1, the frequency of PCR tests and diagnosis code U07.1 was assessed over the whole study period and by month. Characteristics were summarized with frequency distributions for categorical variables and mean (standard deviation [SD]) and median for continuous variables. The closest record of a diagnosis code within 2 days of a PCR test result was validated against the latter. If a patient had multiple PCR tests, the first one during the period of interest was used. Patients were classified as true positive (TP) if they had code U07.1 and were positive by PCR; false positive (FP) if they had code U07.1 but were negative by PCR; true negative (TN) if they lacked code U07.1 and were negative by PCR; and false negative (FN) if they lacked code U07.1 but were positive by PCR. Sensitivity was calculated as $TP/(TP+FN)$; specificity as $TN/(TN+FP)$; positive predictive value (PPV) as $TP/(TP+FP)$; and negative predictive value (NPV) as $TN/(TN+FN)$.

For Part 2, baseline characteristics of patients hospitalized with COVID-19 were summarized. To validate the EMR-based case definition, 300 patients with severe COVID-19 illness were classified as TP or FP and 100 patients without severe illness were classified as TN or FN based on chart review (the gold standard), and PPV and NPV were calculated.

RESULTS

In Part 1, 354,124 patients met the inclusion criteria and had ≥ 1 COVID-19 diagnostic measure from March 1 to November 19, 2020. Part 2 of this study analyzed 3,580 of these patients who had ≥ 1 positive PCR test and were hospitalized with COVID-19 (**Fig. S1**).

Distribution of COVID-19 diagnostic measures

Overall, 94.6% of patients had ≥ 1 PCR test and 6.5% had ≥ 1 positive PCR test (**Table S1**). Of the 726,652 PCR tests administered, 5.9% were positive. From March to November 2020, 11.2% of patients had ≥ 1 diagnosis code U07.1; the mean (SD) number of codes per patient was 3.0 (5.4).

Trends in COVID-19 diagnostic tests and coding

The number of patients per month receiving a PCR test increased from 11,412 in March to 71,637 in August and remained elevated through November (**Fig. 1a**). The number of patients per month with diagnosis code U07.1 increased from 906 in March to 10,700 in April but declined between May (9,779) and June (5,036), remaining stable thereafter.

Positive PCR tests and diagnosis code U07.1 peaked in April before declining in May (**Fig. 1b**) and remained relatively stable from June through November. There were consistently more patients with diagnosis code U07.1 than with a positive PCR test from April to November.

Validation of the diagnosis code against the PCR test

The sensitivity of diagnosis code U07.1 against the PCR test was 53.8% from March to November and varied over the study period, being highest in April to June ($\geq 60\%$) and lowest in October and November ($\sim 20\%$) (**Table 1**). The specificity of the diagnosis code remained high ($>90\%$) during the study period and was 96.6% overall. The PPV of the diagnosis code was

47.4% over the study period and was highest (>70%) in March and April. The NPV of the diagnosis code was high (>88%) across all months and was 97.3% overall.

Characteristics of the population with severe COVID-19 illness

On the index date, patients hospitalized with COVID-19 had a mean age of 59 years; 57% were male, and a large proportion (42%) were overweight or obese (body mass index ≥ 25 kg/m²) (**Table 2**). Common comorbidities included hypertension (44%), chronic kidney disease/dialysis (31%), and diabetes mellitus (30%). The first positive PCR test result for most patients (64%) was on the index date and was performed in an inpatient setting (57%). The main symptoms experienced by patients were shortness of breath (33%), cough (26%), and fever (25%).

Frequency of severe or critical COVID-19 illness endpoints in patients hospitalized with COVID-19

Among the 3,580 patients hospitalized with COVID-19, the mean (SD) duration of hospitalization was 13.6 (17.6) days (**Table 3**); 94.9% and 75.3% met the Pfizer–BioNTech COVID-19 vaccine trial and WHO definitions of severe COVID-19, respectively, whereas 40.6% met the WHO definition of critical illness (**Tables 3** and **S2**). Among 300 patients with severe COVID-19 illness per the vaccine trial definition in codified data who were randomly selected for chart review, 89.0% had SpO₂ $\leq 93\%$ and 50.7% had respiratory frequency >30 breaths/min; 59.3% had significant acute renal, hepatic, or neurologic dysfunction; 44.7% showed evidence of shock; 43.0% were admitted to the ICU; 26.7% had respiratory failure; and the rate of in-facility death was 11.7%.

Among patients meeting the WHO definition of severe COVID-19 illness,¹⁵ 54.0% had SpO₂ <90% and 45.5% had a respiratory frequency >30 breaths/min, with 47.6% showing signs of

severe respiratory distress (**Table S2**). Among patients meeting the WHO definition of critical COVID-19 illness, 27.8% required life-sustaining treatment including vasopressors (25.5%) and noninvasive or mechanical ventilation (37.5%).

Validation of EMR-based case definitions for severe or critical COVID-19 illness endpoints by medical chart review

Severe COVID-19 illness endpoints in the EMR-based case definition had a high PPV (>95%), as validated through chart review (**Table 4**). The PPV of severe COVID-19 illness endpoints was high from March through June (94%–100%), but declined to <90% in subsequent months, which may have been due to the limited sample size (**Fig. S2**).

Severe COVID-19 illness endpoints defined by the WHO had a PPV of 78.5% in the 237 patients meeting this definition (**Table S3**). The NPV of the 100 patients who did not meet the vaccine trial definition of severe COVID-19 illness or the WHO definition of severe or critical COVID-19 illness in codified data was 75.0% and 91.7%, respectively (**Table S4**).

DISCUSSION

Correct identification of positive cases and classification of COVID-19 illness severity are essential for assessing the efficacy and safety of vaccines and treatments in real-world settings. Observational studies are particularly important in COVID-19 research given that vaccines and treatments are initially authorized under EUA, and require supporting data on long-term efficacy and safety. Specifically, accurate COVID-19 endpoint case definitions that can be applied to observational longitudinal data are important from a methodologic standpoint. This was addressed in the present study by validating clinical trial-based COVID-19 diagnosis and severe COVID-19 illness endpoints based on the use of the ICD-10-CM code U07.1 for COVID-19

against the PCR test in a US patient cohort between March and November 2020. The results showed that the diagnosis code had low sensitivity and PPV but high specificity against the PCR test. Meanwhile, the EMR-based case definition for severe COVID-19 illness endpoints had a high PPV among patients hospitalized with COVID-19.

Following its introduction on April 1, 2020 by the Centers for Disease Control and Prevention,¹⁶ the ICD-10-CM code U07.1 for a confirmed diagnosis of COVID-19 was widely adopted by hospitals in the US.¹⁷ An analysis of inpatient hospital data from April to May 2020 in the Premier Healthcare Database reported a high sensitivity (98%) and specificity (99%) for the code compared with PCR, with a PPV of 92% and NPV of 99.97%.¹⁸ The lower sensitivity of the code compared with the PCR test (54%) in the present study may be explained by the inclusion in the MGB Data Mart of data from both inpatient and outpatient settings and not just hospitalized patients, who likely exhibit more severe and recognizable symptoms that facilitate accurate diagnosis. Another study that examined the concordance between coding with U07.1 and the PCR test using MGB data found that among hospitalized patients discharged between April and July 2020, the code had an overall sensitivity of 49%,¹⁹ which is closer to the value in the present study; however, the PPV of 90% was much higher than what we observed (23%–79%) during the same period. The impact of setting on the validity of diagnosis code U07.1 is supported by a retrospective analysis of Department of Veterans Affairs medical records from inpatient, outpatient, and emergency/urgent care settings from April 2020 to March 2021, which found that the PPV of code U07.1 was lower in outpatients settings (78% vs. 84% overall).²⁰ The present study's results suggest that the diagnosis code may have been used in cases where COVID-19–like symptoms were present but a definitive diagnosis was not confirmed by testing, or that the diagnosis code was used for billing or ordering laboratory tests. This practice may be attributable

to evolving medical guidelines during COVID-19 case surges, shortage of PCR tests, or changes in hospital procedures related to PCR test administration. The reliability of code U07.1 may be improved by its combined use with the new COVID-19–related diagnosis and procedure codes released in January 2021²¹; for example, in the study by Lynch et al.,²⁰ the major contributor to the FP rate for U07.1 was a history of COVID-19, for which there is now a separate code. PPV is also impacted by disease prevalence, with PPV being higher when disease prevalence is higher. As COVID-19 prevalence fluctuates over time and by geography, PPV will also vary accordingly, which may explain the variation in PPV results across studies conducted in different time periods and different populations.

The FDA recommends that COVID-19 vaccine trials evaluate severe COVID-19 illness as an endpoint.² However, the definition of severe COVID-19 illness has evolved over time and there are now multiple and overlapping definitions.^{2,22-24} In the present study, 95% and 75% of patients hospitalized with COVID-19 met the BNT162b2 vaccine trial and WHO definitions of severe illness, respectively, and 41% met the WHO definition of critical illness. We assessed the validity of severe COVID-19 illness endpoints through a review of medical charts of a random sample of patients hospitalized with COVID-19. The PPV of each individual endpoint of severe illness derived from the clinical trial and included in the EMR-based case definition was >95%. Endpoints for severe COVID-19 illness according to the WHO had a comparatively lower PPV (79%), but the PPV of most WHO-defined critical COVID-19 illness endpoints was high (>91%). Thus, the BNT162b2 trial severe COVID-19 illness and WHO critical COVID-19 illness endpoints can be reliably used to identify severe cases of COVID-19 from EMR databases, which may facilitate the interpretation of COVID-19 patient data such as long-term outcomes in future studies. This is supported by recent studies that used real-world data from the

US population to confirm the effectiveness of the BNT162b2 vaccine in preventing hospitalization and severe illness.²⁵⁻²⁹

This study had certain limitations. First data from March to November 2020 were used because the study was initiated in late 2020 and because of Institutional Review Board approval timelines, lag time for data availability, and the time required to conduct chart review, more recent data were not included. The study period covers the time prior to the availability of COVID-19 vaccines and a portion of the SARS-CoV-2 Alpha variant period. Therefore, if the patterns of use of diagnosis code U07.1 and physician recordings of COVID-19 illness signs and symptoms changed over the course of the pandemic and have been affected by vaccination status and surges of the Delta and Omicron variants, it is possible that our results cannot be extrapolated to the post-vaccination era. Second, as MGB is a regional data source centered in the Boston metropolitan area and consists of teaching hospitals affiliated with Harvard Medical School, the results may not be generalizable to all patients with COVID-19 in the US. Third, the database may have contained miscoded or incomplete information such as patient care received outside of the MGB network, which may have resulted in underreporting. Fourth, the sample sizes for the number of ICD-10-CM codes (e.g., in March or October 2020) were limited in certain analyses. Finally, validation of the EMR-based case definition for severe COVID-19 illness endpoints was based on PPV and NPV. However, PPV decreases with decreasing COVID-19 prevalence because the rate of FP increases. Thus, the fluctuation of COVID-19 prevalence could complicate the interpretation of the summary PPV estimate generated in this study as well as its generalizability to different geographic locations or time periods with surges in COVID-19 variants. Sensitivity and specificity are important measures that are not impacted by COVID-19 prevalence; however, given the logistics of chart review, their calculation would

have required a large random sample of patients to ensure an adequate number of true cases, rendering sampling from true positive and negative COVID-19 severe cases infeasible.

CONCLUSION

The ICD-10-CM diagnosis code for COVID-19 (U07.1) did not have adequate sensitivity to identify COVID-19 patients compared with the PCR test; therefore, positive cases recorded using the diagnosis code may require confirmation by PCR testing. On the other hand, the EMR-based case definition of severe COVID-19 illness endpoints showed high positive predictive performance and may be applicable to analyses of real-world COVID-19 patient datasets.

ACKNOWLEDGMENTS

This study was funded by Pfizer, Inc., New York, NY, USA. The study sponsor was involved in several aspects of the research including study design, data interpretation, manuscript writing, and decision to submit the manuscript for publication. The authors thank Ann Madsen, PhD (a former employee of Pfizer) for her methodologic input in the study. Medical writing assistance was provided by Janice Imai of Analysis Group, Montreal, Canada (funding for medical writing was provided by Pfizer, Inc., New York, NY, USA).

CONFLICT OF INTEREST

The authors have the following conflicts of interest to declare: Mei Sheng Duh, Catherine Nguyen, Rose Chang, Maral DerSarkissian, Azeem Banatwala, Louise H. Yu, Bruce E. Stangle, and Pierre Y. Cremieux are employees of Analysis Group, Inc., which received research funding from Pfizer for this study. Heather Rubino and Francesca Kolitsopoulos are full-time employees and stock shareholders of Pfizer. Christopher Herrick, Yichuan Grace Hsieh, Gregory Belsky,

Marykate E. Murphy, Janet Boyle-Kelly, Andrew Cagan, and Shawn N. Murphy have nothing to disclose.

ETHICS STATEMENT

This study was approved by the MGB Institutional Review Board.

PATIENT CONSENT STATEMENT

Patient consent was not required as patient data were de-identified.

REFERENCES

1. World Health Organization (WHO). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. (<https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>).
2. Schoeni RF, Wiemers EE, Seltzer JA, Langa KM. Association between risk factors for complications from COVID-19, perceived chances of infection and complications, and protective behavior in the US. *JAMA Netw Open* 2021;4(3):e213984.
3. Center for Disease Control and Prevention (CDC). People with certain medical conditions. (<https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html>).
4. Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K, Danon L. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ* 2021;372:n579. DOI: doi:10.1136/bmj.n579.
5. Polack FP, Thomas SJ, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med* 2020;383:2603–15. DOI: <https://doi.org/10.1056/NEJMoa2034577>.
6. Kirby T. New variant of SARS-CoV-2 in UK causes surge of COVID-19. *Lancet Respir Med* 2021;9(2):e20-1.
7. Walsh EE, Frenck RW, Jr., Falsey AR, et al. Safety and immunogenicity of two RNA-based Covid-19 vaccine candidates. *N Engl J Med* 2020;383(25):2439-2450. DOI: 10.1056/NEJMoa2027906.
8. United States Centers for Disease Control and Prevention. COVID Data Tracker – COVID-19 Vaccinations in the United States, https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-total-admin-rate-total; [accessed 6 September 2022].
9. Danza P, Koo TH, Haddix M, et al SARS-CoV-2 infection and hospitalization among adults aged ≥18 years, by vaccination status, before and during SARS-CoV-2 B.1.1.529 (omicron) variant predominance — Los Angeles County, California, November 7, 2021–January 8, 2022. *MMWR Morb Mortal Wkly Rep* 2022;71:177–181. DOI: 10.15585/mmwr.mm7105e1.
10. Johnson AG, Amin AB, Ali AR, et al. COVID-19 incidence and death rates among unvaccinated and fully vaccinated adults with and without booster doses during periods of Delta and Omicron variant emergence — 25 U.S. jurisdictions, April 4–December 25, 2021. *MMWR Morb Mortal Wkly Rep* 2022;71:1–7. DOI: <http://dx.doi.org/10.15585/mmwr.mm7104e2>.
11. Sun Q, Qiu H, Huang M, et al. Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province. *Ann Intensive Care* 2020;10:33. DOI: <https://doi.org/10.1186/s13613-020-00650-2>.
12. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* 2020;25:2000045. DOI: <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>.
13. Mass General Brigham Research Information Science and Computing, <https://rc.partners.org/about/who-we-are-risc/research-patient-data-registry>; [accessed 19 January 2022].
14. Shimabukuro T. National Center for Immunization & Respiratory Disease: Enhanced safety monitoring for COVID-19 vaccines in early phase vaccination. (<https://www.cdc.gov/vaccines/acip/meetings/downloads/slides-2020-09/COVID-03-Shimabukuro.pdf>).
15. WHO. COVID-19 Therapeutic Trial Synopsis. (https://www.who.int/blueprint/priority-diseases/key-action/COVID-19_Treatment_Trial_Design_Master_Protocol_synopsis_Final_18022020.pdf2020).
16. United States Centers for Disease Control and Prevention. ICD-10-CM Official Coding and Reporting Guidelines April 1, 2020 through September 30, 2020,

- <https://www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf>; 2020 [accessed 31 December 2021].
17. Yang KC, Lee B, Ahn YY, et al. Use of and comorbidities associated with diagnostic codes for COVID-19 in US health insurance claims. *JAMA Netw Open* 2021;4:e2124643. DOI: <https://doi.org/10.1001/jamanetworkopen.2021.24643>.
 18. Kadri SS, Gundrum J, Warner S, et al. Uptake and accuracy of the diagnosis code for COVID-19 among US hospitalizations. *JAMA* 2020;324:2553–4. DOI: <https://doi.org/10.1001/jama.2020.20323>.
 19. Bhatt AS, McElrath EE, Claggett BL, et al. Accuracy of ICD-10 diagnostic codes to identify COVID-19 among hospitalized patients. *J Gen Intern Med* 2021;36:2532–5. DOI: <https://doi.org/10.1007/s11606-021-06936-w>.
 20. Lynch KE, Viernes B, Gatsby E, et al. Positive predictive value of COVID-19 ICD-10 diagnosis codes across calendar time and clinical setting. *Clin Epidemiol* 2021;13:1011–8. DOI: <https://doi.org/10.2147/CLEP.S335621>.
 21. United States Centers for Disease Control and Prevention. ICD-10-CM Official Guidelines for Coding and Reporting FY 2021 – UPDATED January 1, 2021 (October 1, 2020–September 30, 2021), <https://www.cdc.gov/nchs/data/icd/ICD-10cmguidelines-FY2021-COVID-update-January-2021-508.pdf>; 1 January 2021 [accessed 2 January 2022].
 22. Wu Z, McGoogan JM. 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* 2020;323:1239–42. DOI: <https://doi.org/10.1001/jama.2020.2648>.
 23. National Institutes of Health. COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines, <https://www.covid19treatmentguidelines.nih.gov/>; 19 October 2021 [accessed 30 December 2021].
 24. Zhou YH, Li H, Qin YY, et al. Predictive factors of progression to severe COVID-19. *Open Med (Wars)* 2020;15:805–14. DOI: <https://doi.org/10.1515/med-2020-0184>.
 25. Cohn BA, Cirillo PM, Murphy CC, et al. SARS-CoV-2 vaccine protection and deaths among US veterans during 2021. *Science* 2021:eabm0620. DOI: <https://doi.org/10.1126/science.abm0620>.
 26. Lutrick K, Rivers P, Yoo YM, et al. Interim estimate of vaccine effectiveness of BNT162b2 (Pfizer–BioNTech) vaccine in preventing SARS-CoV-2 infection among adolescents aged 12–17 Years — Arizona, July–December 2021. *MMWR Morb Mortal Wkly Rep* 2021;70:1761–5. DOI: <http://dx.doi.org/10.15585/mmwr.mm705152a2>.
 27. Nanduri S, Pilishvili T, Derado G, et al. Effectiveness of Pfizer–BioNTech and Moderna vaccines in preventing SARS-CoV-2 infection among nursing home residents before and during widespread circulation of the SARS-CoV-2 B.1.617.2 (Delta) variant — National Healthcare Safety Network, March 1–August 1, 2021. *MMWR Morb Mortal Wkly Rep* 2021;70:1163–6. DOI: <http://dx.doi.org/10.15585/mmwr.mm7034e3external>.
 28. Olson SM, Newhams MM, Halasa NB, et al. Effectiveness of Pfizer–BioNTech mRNA vaccination against COVID-19 hospitalization among persons aged 12–18 years — United States, June–September 2021. *MMWR Morb Mortal Wkly Rep* 2021;70:1483–8. DOI: <http://dx.doi.org/10.15585/mmwr.mm7042e1external>.
 29. Thompson MG, Burgess JL, Naleway AL, et al. Prevention and Attenuation of Covid-19 with the BNT162b2 and mRNA-1273 Vaccines. *N Engl J Med* 2021;385:320–9. DOI: <https://doi.org/10.1056/NEJMoa2107058>.

Figure 1. Number of patients with each COVID-19 diagnostic measure and positive PCR test or ICD-10-CM code by month. **A.** Number of patients with PCR test or ICD-10-CM code by month. **B.** Number of patients with positive COVID-19 PCR test or ICD-10-CM code by month. Data for November were available only up to November 19, 2020.

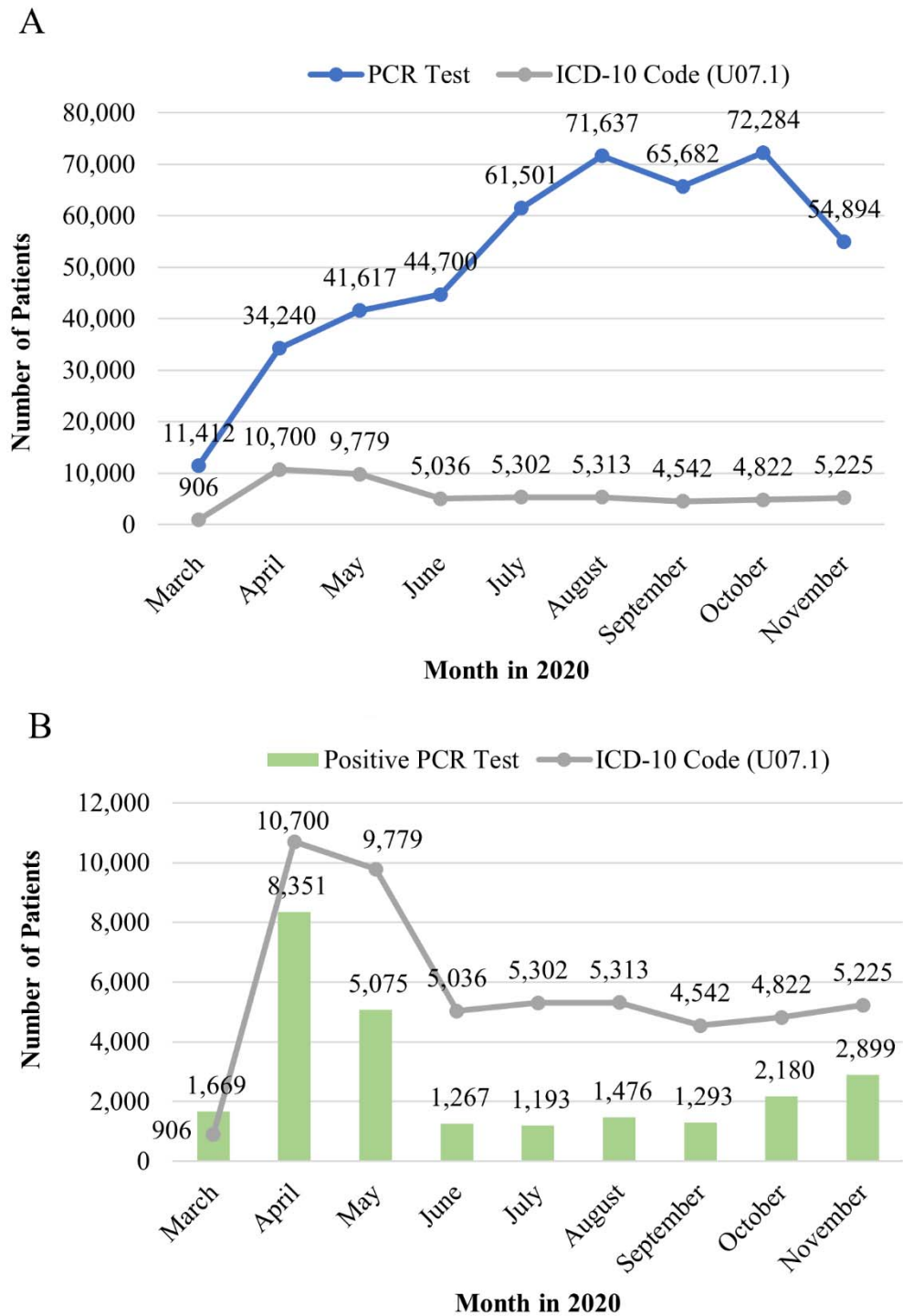


Table 1. Validation of COVID-19 ICD-10-CM diagnosis code against PCR test

	Sensitivity [†]			Specificity [†]			PPV [†]			NPV [†]		
	TP	TP + FN	Sensitivity % (95% CI) [‡]	TN	TN + FP	Specificity % (95% CI) [‡]	TP	TP + FP	PPV % (95% CI) [‡]	TN	TN + FN	NPV % (95% CI) [‡]
Overall [†]	9,819	18,237	53.8 (53.1, 54.6)	305,893	316,806	96.6 (96.5, 96.6)	9,819	20,732	47.4 (46.7, 48.0)	305,893	314,311	97.3 (97.3, 97.4)
March 2020	374	1,648	22.7 (20.7, 24.8)	9,632	9,764	98.6 (98.4, 98.9)	374	506	73.9 (69.9, 77.7)	9,632	10,906	88.3 (87.7, 88.9)
April 2020	5,627	8,009	70.3 (69.2, 71.3)	24,743	26,231	94.3 (94.0, 94.6)	5,627	7,115	79.1 (78.1, 80.0)	24,743	27,125	91.2 (90.9, 91.6)
May 2020	2,990	4,652	64.3 (62.9, 65.7)	34,204	36,965	92.5 (92.3, 92.8)	2,990	5,751	52.0 (50.7, 53.3)	34,204	35,866	95.4 (95.1, 95.6)
June 2020	787	1,166	67.5 (64.7, 70.2)	41,825	43,534	96.1 (95.9, 96.3)	787	2,496	31.5 (29.7, 33.4)	41,825	42,204	99.1 (99.0, 99.2)
July 2020	619	1,114	55.6 (52.6, 58.5)	58,341	60,387	96.6 (96.5, 96.8)	619	2,665	23.2 (21.6, 24.9)	58,341	58,836	99.2 (99.1, 99.2)
August 2020	740	1,386	53.4 (50.7, 56.0)	67,996	70,251	96.8 (96.7, 96.9)	740	2,995	24.7 (23.2, 26.3)	67,996	68,642	99.1 (99.0, 99.1)
September 2020	704	1,191	59.1 (56.3, 61.9)	62,581	64,491	97.0 (96.9, 97.2)	704	2,614	26.9 (25.2, 28.7)	62,581	63,068	99.2 (99.2, 99.3)
October 2020	389	1,999	19.5 (17.7, 21.3)	68,148	70,285	97.0 (96.8, 97.1)	389	2,526	15.4 (14.0, 16.9)	68,148	69,758	97.7 (97.6, 97.8)
November 2020 [§]	546	2,636	20.7 (19.2, 22.3)	50,391	52,258	96.4 (96.3, 96.6)	546	2,413	22.6 (21.0, 24.4)	50,391	52,481	96.0 (95.8, 96.2)

[†]Sensitivity was calculated as TP/ (TP + FN); specificity was calculated as TN/ (TN + FP); PPV was calculated as TP/ (TP + FP); and NPV was calculated as FN/ (TN + FN). TP was the presence of the ICD-10-CM code and positive PCR test; TP + FN was a positive PCR test; TN was the absence of the ICD-10-CM code and negative PCR test; TN + FP was a negative PCR test; TP + FP was the presence of the ICD-10-CM code; and TN + FN was the absence of the ICD-10-CM code.

[‡]Exact CIs are presented for each validation measure.

[¶]All patients with a PCR test from March to November 2020 were included in the analysis. The presence and absence of the ICD-10-CM diagnosis code (U07.1) within 2 days of a PCR test was categorized as “test positive” and “test negative”, respectively. If a patient had multiple PCR tests, the first PCR test in the period from March to November 2020 was used in the overall analysis, and the first PCR test in the month was used in the by-month analysis.

[§]Data for November were available only up to November 19, 2020.

Abbreviations: CI, confidence interval; COVID-19, coronavirus disease 2019; FN, false negative; FP, false positive; ICD-10-CM, International Classification of Diseases 10th Revision, Clinical Modification; NPV, negative predictive value; PCR, polymerase chain reaction; PPV, positive predictive value; TN, true negative; TP, true positive.

Table 2. Baseline characteristics of patients hospitalized with COVID-19

Characteristic [†]	All patients hospitalized with COVID-19 N = 3,580
Demographic characteristics	
Age on index date, years	
Mean ± SD	59.0 ± 16.3
Median (IQR)	60.0 (48.0, 72.0)
Age on index date, years, n (%)	
18–29	177 (4.9)
30–39	351 (9.8)
40–49	450 (12.6)
50–59	711 (19.9)
60–69	799 (22.3)
70–79	717 (20.0)
80–86	375 (10.5)
Sex, n (%)	
Male	2,027 (56.6)
Female	1,553 (43.4)
Race, n (%)	
White	1,704 (47.6)
Black or African American	631 (17.6)
Asian	116 (3.2)
American Indian/Alaska Native	9 (0.3)
Native Hawaiian/Pacific Island	7 (0.2)
Other [‡]	896 (25.0)
Unknown	217 (6.1)
Smoking status, n (%) [¶]	
Never smoker	1,418 (39.6)
Former smoker	372 (10.4)
Current smoker	293 (8.2)
Unknown	1,497 (41.8)
Clinical characteristics	
BMI, kg/m ² , n (%) [¶]	
Underweight (<18.5)	31 (0.9)
Normal weight (18.5–24.9)	353 (9.9)
Overweight (25–29.9)	594 (16.6)
Obese (≥30)	909 (25.4)
Unknown	1,693 (47.3)
Quan-CCI [§]	
Mean ± SD	1.3 ± 1.9
Median (IQR)	0.0 (0.0, 2.0)
Comorbidities, n (%) [□]	
Hypertension	1,575 (44.0)
Chronic kidney disease/dialysis	1,099 (30.7)

Diabetes mellitus	1,064 (29.7)
Hyperlipidemia	985 (27.5)
Chronic obstructive pulmonary disease/interstitial lung disease	793 (22.2)
Coronary artery disease	655 (18.3)
Asthma	333 (9.3)
Bleeding diathesis or condition associated with prolonged bleeding	324 (9.1)
Cancer	306 (8.5)
Liver disease	260 (7.3)
Venous thromboembolism	260 (7.3)
Autoimmune disease	195 (5.4)
HIV/HBV/HCV	108 (3.0)
Other immune deficiencies	103 (2.9)
Solid organ transplant	4 (0.1)
Concomitant medications, n (%) [#]	
Antibiotics	1,925 (53.8)
NSAIDs	1,114 (31.1)
Statins	1,192 (33.3)
Corticosteroids	729 (20.4)
Previous vaccine anaphylaxis, n (%) [□]	4 (0.1)
Disease characteristics	
Vital signs on index date, n (%) ^{††}	
Patients with temperature available	3,556 (99.3)
<37.5°C (normal)	1,705 (47.9)
37.5°C–38.0°C	639 (18.0)
38.1°C–39.0°C	807 (22.7)
>39.0°C	405 (11.4)
Patients with heart rate available	3,567 (99.6)
<125 beats/min (normal)	3,147 (88.2)
≥125 beats/min	420 (11.8)
Patients with respiratory frequency available	3,515 (98.2)
<30 breaths/min (normal)	2,402 (68.3)
≥30 breaths/min	1,113 (31.7)
Patients with SpO ₂ available	3,564 (99.6)
≤93%	2,003 (56.2)
>93% (normal)	1,561 (43.8)
Patients with SBP available	3,559 (99.4)
<90 mm Hg	376 (10.6)
≥90 mm Hg	3,183 (89.4)
Time from first positive PCR test to index date, n (%)	
7–14 days prior to index date	222 (6.2)
≤7 days prior to index date	640 (17.9)
On index date	2,291 (64.0)
≤2 days after index date	427 (11.9)
Care setting at COVID-19 diagnosis (i.e., positive PCR test), n (%) ^{‡‡}	
Inpatient	2,036 (56.9)
Emergency room	191 (5.3)

Outpatient	118 (3.3)
Unknown	1,235 (34.5)
Symptoms reported on index date, n (%) ^{¶¶}	
Shortness of breath	1,204 (33.6)
Cough	932 (26.0)
Fever	889 (24.8)
Diarrhea	236 (6.6)
Myalgia	127 (3.5)
Headache	141 (3.9)
Sore throat	64 (1.8)
Rhinorrhea	15 (0.4)

Data are shown as n (%) unless otherwise indicated.

[†]Index date (period up to 3 months prior to the date of first hospitalization with COVID-19).

[‡]Includes “Two or more races” or “Other race”.

[¶]Status on or on the date closest (in the 12 months prior) to the index date.

[§]Assessed during baseline period and on the index date using diagnosis codes. Reference: Quan et al. (2005). Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10-CM administrative data. *Medical Care*, 43(11): 1130–39.

[□] Assessed during baseline period and on the index date using diagnosis and procedure codes.

[#] Assessed during baseline period and on the index date.

^{††} Assessed on the index date. If multiple test results were available, the most severe result for each test was used.

^{‡‡} For first positive PCR test within 14 days before or 2 days after hospitalization.

^{¶¶} Assessed on the index date using diagnosis codes. Multiple symptoms may be reported for each patient.

Abbreviations: BMI, body mass index; CCI, Charlson Comorbidity Index; COVID-19, coronavirus disease 2019; FEU, fibrinogen equivalent units; HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; IQR, interquartile range; NSAID, nonsteroidal anti-inflammatory drug; PCR, polymerase chain reaction; SBP, systolic blood pressure; SD, standard deviation; SpO₂, oxygen saturation.

Table 3. Frequency of severe COVID-19 illness endpoints in patients hospitalized with COVID-19

Endpoint	All patients hospitalized with COVID-19 [†] N = 3,580	Random sample of patients meeting vaccine trial definition of severe COVID-19 illness [‡] N = 300
Duration of COVID-19–associated hospitalization, days, mean ± SD [median]	13.6 ± 17.6 [8.0]	14.7 ± 18.3 [8.0]
Patients meeting vaccine trial definition of severe COVID-19 illness[¶], n (%)	3,398 (94.9)	300 (100.0)
Clinical signs at rest, n (%)[§]	3,163 (88.4)	279 (93.0)
Patients with SpO ₂ available	3,578 (99.9)	300 (100.0)
SpO ₂ ≤93%	3,021 (84.4)	267 (89.0)
Patients with PaO ₂ /FiO ₂ available [□]	639 (17.8)	51 (17.0)
PaO ₂ /FiO ₂ <300 mm Hg	325 (50.9)	27 (52.9)
Patients with respiratory frequency available	3,573 (99.8)	298 (99.3)
Respiratory frequency ≥30 breaths/min	1,808 (50.6)	151 (50.7)
Patients with heart rate available	3,580 (100.0)	300 (100.0)
Heart rate ≥125 beats/min	1,015 (28.4)	86 (28.7)
Significant acute renal or hepatic, or neurologic dysfunction, n (%)[#]	1,969 (55.0)	178 (59.3)
Significant acute renal dysfunction	1,351 (37.7)	131 (43.7)
Acute kidney failure	984 (27.5)	87 (29.0)
Patients with serum creatinine available	970 (27.1)	100 (33.3)
Increase in serum creatinine by ≥0.3 mg/dl within 48 h	826 (85.2)	83 (83.0)
Increase in serum creatinine to ≥1.5× baseline within previous 7 days	781 (80.5)	75 (75.0)
Significant acute hepatic dysfunction	1,170 (32.7)	97 (32.3)
Patients with AST or ALT available ⁴	3,399 (94.9)	288 (96.0)
AST or ALT >3.0× ULN	953 (28.0)	79 (27.4)
Patients with total blood bilirubin, GGT, or ALP available	3,403 (95.1)	289 (96.3)
Blood bilirubin or GGT or ALP >2.0× ULN	562 (16.5)	60 (20.8)
Elevation of transaminase and lactic acid dehydrogenase	393 (11.0)	21 (7.0)
Liver disease	38 (1.1)	5 (1.7)
Hepatic failure	30 (0.8)	4 (1.3)
Hepatomegaly	8 (0.2)	1 (0.3)

Inflammatory liver disease	18 (0.5)	1 (0.3)
Toxic liver disease	16 (0.4)	1 (0.3)
Jaundice	8 (0.2)	1 (0.3)
Central hemorrhagic necrosis of liver	0 (0.0)	0 (0.0)
Significant neurologic dysfunction	287 (8.0)	25 (8.3)
Convulsions/seizures	150 (4.2)	13 (4.3)
Cerebrovascular non-hemorrhagic stroke	132 (3.7)	10 (3.3)
Multiple sclerosis	19 (0.5)	1 (0.3)
Encephalitis/encephalomyelitis	9 (0.3)	0 (0.0)
Other acute demyelinating diseases	8 (0.2)	0 (0.0)
Bell's palsy	7 (0.2)	2 (0.7)
Optic neuritis	3 (0.1)	0 (0.0)
Guillain–Barre syndrome	2 (0.1)	1 (0.3)
Aseptic meningitis	1 (0.0)	0 (0.0)
Transverse myelitis	1 (0.0)	0 (0.0)
Evidence of shock, n (%)	1,444 (40.3)	134 (44.7)
Patients with SBP available [§]	3,579 (100.0)	300 (100.0)
SBP <90 mm Hg	1,187 (33.2)	114 (38.0)
Vasopressors ^{††}	914 (25.5)	81 (27.0)
Admission to ICU^{‡‡}, n (%)	1,435 (40.1)	129 (43.0)
Respiratory failure^{¶¶}, n (%)	927 (25.9)	80 (26.7)
Invasive ventilation	738 (20.6)	66 (22.0)
Noninvasive ventilation	605 (16.9)	49 (16.3)
High-flow oxygen therapy	448 (12.5)	34 (11.3)
ECMO	34 (0.9)	0 (0.0)
In-facility death, n (%)	368 (10.3)	35 (11.7)

[†]For all clinical indices except for PaO₂/FiO₂, blood pressure, and estimated glomerular filtration rate/creatinine clearance, if the patient had multiple measurements on the same day the most severe result was used.

[‡]A random sampling without replacement was used to select 300 patients who met the Pfizer/BioNTech COVID-19 vaccine trial definition of severe COVID-19 illness.

[¶]Patients who met the Pfizer/BioNTech COVID-19 vaccine trial definition of severe COVID-19 illness had at least one of the listed endpoints during hospitalization with COVID-19.

[§]Endpoints were identified from clinical sign-related or laboratory data curated by Mass General Brigham.

□ PaO₂/FiO₂ (i.e., P/F ratio) was calculated by first identifying separate measures for PaO₂ and FiO₂ in the curated laboratory data. FiO₂ values in l/min were converted to percentages using the equation FiO₂ % = (21% + 4% for each l/min of O₂). If a patient had multiple measurements on the same day, the daily average was used. P/F ratio was then calculated by dividing PaO₂ in mm Hg by FiO₂ %.

#Definitions of acute renal, hepatic, or neurologic dysfunction were based on the Brighton Collaboration's definition of a list of COVID-19 adverse events of special interest.

††Vasopressors were identified based on string searches of medication names including dopamine, epinephrine, norepinephrine, and vasopressin.

‡‡Identified from ICU admission records curated by Mass General Brigham.

¶¶Procedures used in cases of respiratory failure were identified from structured procedure data based on procedure codes and from respiratory therapy-related procedure data curated by Mass General Brigham.

Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; COVID-19, coronavirus disease 2019; ECMO, extracorporeal membrane oxygenation; FiO₂, fraction of inspired oxygen; GGT, gamma-glutamyl transferase; ICU, intensive care unit; PaO₂, partial pressure of oxygen; SBP, systolic blood pressure; SpO₂, oxygen saturation; ULN, upper limit of normal.

Table 4. Validation of severe COVID-19 illness endpoints against medical chart review

Endpoint	PPV ^{†‡}		
	TP	TP + FP	PPV % (95% CI) [¶]
Hospitalization^{§□}			
Admission date	302	320	94.4 (91.3, 96.6)
Discharge date	255	320	79.7 (74.9, 84.0)
Duration of hospitalization	261	320	81.6 (76.9, 85.7)
Patients meeting vaccine trial definition of severe COVID-19 illness^{##††}	282	300	94.0 (90.7, 96.4)
Clinical signs at rest^{**}	267	279	95.7 (92.6, 97.8)
SpO ₂ ≤93%	259	267	97.0 (94.2, 98.7)
PaO ₂ /FiO ₂ <300 mm Hg	27	27	100.0 (87.2, 100.0)
Respiratory frequency ≥30 breaths/min	149	151	98.7 (95.3, 99.8)
Heart rate ≥125 beats/min	82	86	95.3 (88.5, 98.7)
Significant acute renal, hepatic, or neurologic dysfunction^{¶¶}	171	178	96.1 (92.1, 98.4)
Significant acute renal dysfunction	127	131	96.9 (92.4, 99.2)
Acute kidney failure	85	87	97.7 (91.9, 99.7)
Increase in serum creatinine by ≥0.3 mg/dl within 48 h	80	83	96.4 (89.8, 99.2)
Increase in serum creatinine to ≥1.5× baseline occurred within prior 7 days	74	75	98.7 (92.8, 100.0)
Significant acute hepatic dysfunction	95	97	97.9 (92.7, 99.7)
AST or ALT >3.0× ULN	79	79	100.0 (95.4, 100.0)
Blood bilirubin or GGT or ALP >2.0× ULN	58	60	96.7 (88.5, 99.6)
Elevation of transaminase and lactic acid dehydrogenase	21	21	100.0 (83.9, 100.0)
Liver disease	5	5	100.0 (47.8, 100.0)
Hepatic failure	4	4	100.0 (39.8, 100.0)
Hepatomegaly	1	1	100.0 (2.5, 100.0)
Inflammatory liver disease	1	1	100.0 (2.5, 100.0)
Toxic liver disease	1	1	100.0 (2.5, 100.0)
Jaundice	1	1	100.0 (2.5, 100.0)
Significant neurologic dysfunction	22	25	88.0 (68.8, 97.5)
Convulsions/seizures	11	13	84.6 (54.6, 98.1)
Cerebrovascular non-hemorrhagic stroke	9	10	90.0 (55.5, 99.7)
Multiple sclerosis	1	1	100.0 (2.5, 100.0)
Encephalitis/encephalomyelitis	–	–	–

Other acute demyelinating diseases	–	–	–
Bell’s palsy	2	2	100.0 (15.8, 100.0)
Optic neuritis	–	–	–
Guillain-Barre syndrome	1	1	100.0 (2.5, 100.0)
Aseptic meningitis	–	–	–
Transverse myelitis	–	–	–
Evidence of shock	130	134	97.0 (92.5, 99.2)
SBP <90 mm Hg	112	114	98.2 (93.8, 99.8)
Vasopressors ^{§§}	78	81	96.3 (89.6, 99.2)
Admission to ICU^{□□}	125	129	96.9 (92.3, 99.1)
Respiratory failure^{##}	77	80	96.3 (89.4, 99.2)
Invasive ventilation	64	66	97.0 (89.5, 99.6)
Noninvasive ventilation	48	49	98.0 (89.1, 99.9)
High-flow oxygen therapy	32	34	94.1 (80.3, 99.3)
In-facility death	35	35	100.0 (90.0, 100.0)

[†]A total of 300 patients meeting the Pfizer/BioNTech COVID-19 vaccine trial definition of severe COVID-19 illness was selected by random sampling without replacement.

[‡]PPV was calculated as TP / (TP + FP), where TP and FP were patients with severe COVID-19 illness identified by medical chart review and those with the corresponding endpoint determined from codified data, respectively.

[¶]Exact CIs are presented for each validation measure.

[§]Readmissions within 3 days after a prior hospital discharge were combined with the previous hospitalization as a single event. True hospitalization admission/discharge dates were defined as an admission/discharge date from codified data that was within 1 day of the admission/discharge date determined through chart abstraction. True hospitalization duration was defined as a duration of hospitalization calculated based on codified data that was within 2 days of the duration determined through chart abstraction.

[□]Included all hospitalization events determined to be associated with COVID-10 from codified data.

[#]Patients meeting the Pfizer/BioNTech COVID-19 vaccine trial definition of severe COVID-19 illness had at least one of the listed endpoints during hospitalization with COVID-19.

^{††}Patients were defined as FPs if their hospitalizations were not associated with severe COVID-19 illness as determined through chart abstraction.

^{‡‡}Endpoints were identified from clinical sign-related or laboratory data curated by Mass General Brigham.

^{¶¶}Definitions of acute renal, hepatic, or neurologic dysfunction were based on the Brighton Collaboration’s definition of a list of COVID-19 adverse events of special interest.

^{§§}Identified through string searches of medication names including dopamine, epinephrine, norepinephrine, and vasopressin.

^{□□}Identified from ICU admission records curated by Mass General Brigham.

^{##}Procedures used in cases of respiratory failure were identified from structured procedure data based on procedure codes and from respiratory therapy-related procedure data curated by Mass General Brigham.

Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CI, confidence interval; COVID-19, coronavirus disease 2019; FiO₂, fraction of inspired oxygen; FP, false positive; GGT, gamma-glutamyl transferase; ICU, intensive care unit; PaO₂, partial pressure of oxygen; PPV, positive predictive value; SBP, systolic blood pressure; SpO₂, oxygen saturation; TP, true positive; ULN, upper limit of normal.