Accuracy of ICD-10 Diagnostic Codes to Identify COVID-19 Among Hospitalized Patients

BACKGROUND

Health organizations require rapid, reliable access to data on patients infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). On April 1, 2020, the Centers for Disease Control and Prevention (CDC) introduced a new International Classification of Disease, 10th revision (ICD-10) code U07.1 (COVID-19, virus identified),[1] together with specific coding guidance regarding its appropriate use.[2] Government agencies, including the Centers for Medicare and Medicaid Services, use ICD-10 data for ascertainment of COVID-19 hospitalizations,[3] though little is known about the reliability of the U07.1 code in identifying disease. We sought to determine the performance characteristics of the ICD-10 code U07.1 for identification of COVID-19 illness in a large multicenter health system.

METHODS

We identified all inpatient encounters during which ≥1 SARS-CoV-2 reverse transcription polymerase chain reaction (RT-PCR) was performed from April 1 to July 31, 2020, across the Mass General Brigham health system. Patients with ≥1 positive SARS-CoV-2 RT-PCR were denoted to be COVID-19 positive. The agreement between COVID-19 positivity and primary or secondary ICD-10 coding of U07.1 was determined. Performance characteristics (sensitivity, specificity, positive predictive value [PPV], negative predictive value [NPV]) for the ICD-10 code U07.1 were reported overall and across major subgroups. Among COVID-19-positive patients, we performed multivariable logistic regression to identify independent predictors of corresponding ICD-10 coding for COVID-19(sensitivity). The Mass General Brigham Institutional Review Board approved the study protocol. Data management and analysis were performed using STATA (College Station, TX).

RESULTS

There were 22,633 patient encounters with a discharge date between April 1, 2020, and July 31, 2020, in which ≥1 SARS-CoV-2 RT-PCR was obtained during admission. Overall, 66.7%, 25.7%, and 7.7% of encounters had 1, 2, and >2 SARS-CoV-2 RT-PCR test(s) performed, respectively (range 1 to 16). Among these encounters, 2210 (9.8%) were determined to be COVID-19 positive. COVID-19 test–positive patients were older (64±18 vs. 60±19 years) and more likely to be men (51.8% vs. 44.3%), Hispanic (22.3% vs. 8.3%), and Black (21.6% vs. 10.4%) as compared with those with all negative RT-PCR results (P<0.001 for all). ICD-10 diagnostic code U07.1 was coded in 1208 (5.3%) patients. U07.1 had an overall sensitivity of 49.2% (95% confidence interval [CI]: 47.1–51.3%), specificity of 99.4% (95% CI: 99.3–99.5%), PPV of 90.0% (95% CI: 88.2–91.6%), and NPV of 94.8% (95% CI: 94.5–95.0%). Specificity remained high over time: 98.4% in April, 99.0% in May, 99.7% in June, and 99.9% in July 2020, while sensitivity varied and was lowest in July 2020 (27.9%). Sensitivity was lower among those age 0–17 years (14.3%, 95% CI: 1.2–70.1%), although this subset was limited by few encounters (n=175). Similar performance of U07.1 coding was observed across all other major subgroups (Table 1). Earlier months in the pandemic were the only significant independent predictors of higher sensitivity of the ICD-10 diagnostic code U07.1 among COVID-19 test–positive patients (Fig. 1).

DISCUSSION

Uniform administrative coding for research, disease tracking, and quality improvement is appealing given its widespread use and ease of interoperability across health systems. Reliance on these administrative data will likely remain important for prior COVID-19 disease identification, particularly given expanding interest in identifying legacy effects and post-acute sequelae of SARS-CoV-2 (PASC) infection. However, ICD-10 codes in other clinical settings[4] and for COVID-19-related symptoms[5] are known to be subject to misclassification.

We found sensitivity for U07.1 coding among hospitalized patients undergoing SARS-CoV-2 RT-PCR testing was modest, while specificity was high and approached 100% over time. Lack of initial awareness or familiarity with ICD-10 coding for COVID-19, in addition to distinctions between test

Table 1 Performance Characteristics of ICD-10 Code for COVID-19 as Compared with SARS-CoV-2 RT-PCR Positivity

U07.1 ICD-10 code performance characteristics

	# of patient encounters	# of encounters with ≥1 positive SARS-CoV-2 RT-PCR, n (%)	Point estimate, % (95% confidence interval [CI]) Sensitivity	Specificity	PPV	NPV
Overall	22,633	2210 (9.8)	49.2 (47.1–51.3)	99.4 (99.3–99.5)	90.0 (88.2–91.6)	94.8 (94.5–95.0)
Month						
April 2020	3827	1045 (27.3)	45.3 (42.3–48.8)	98.4 (97.9–98.8)	91.5 (88.7–93.6)	82.7 (81.4–84.0)
May 2020	5982	726 (12.1)	60.9 (57.3–64.4)	99.0 (98.7–99.2)	89.3 (86.2–91.7)	94.8 (94.2–95.4)
June 2020	6108	292 (4.8)	44.9 (39.2–50.6)	99.7 (99.5–99.8)	88.5 (82.2–92.8)	97.3 (96.9–97.7)
July 2020	6716	147 (2.2)	27.9 (21.2–35.7)	99.9 (99.8–99.9)	85.4 (72.0–93.0)	98.4 (98.1–98.7)
Age (years)						
0–17	175	7 (4.0)	14.3 (1.2–70.1)	100.0 (–)	100.0 (–)	96.6 (92.5–98.5)
18–34	3116	171 (5.5)	42.7 (35.4–50.3)	99.7 (99.5–99.9)	90.1 (81.3–95.0)	96.8 (96.1–97.3)
35–64	8917	877 (9.8)	49.0 (45.7–52.3)	99.2 (99.0–99.4)	87.6 (84.3–90.2)	94.7 (94.2–95.2)
≥65	10,425	1155 (11.1)	50.5 (47.6–53.4)	99.4 (99.3–99.6)	91.8 (89.4–93.7)	94.2 (93.7–94.6)
Gender						
Men	10,191	1144 (11.2)	48.9 (46.0–51.8)	99.2 (99.0–99.3)	88.2 (85.4–90.5)	93.9 (93.4–94.3)
Women	12,422	1066 (8.5)	49.5 (46.5–52.5)	99.6 (99.5–99.7)	92.0 (89.5–93.9)	95.5 (95.1–95.8)
Race						
White	16,858	1170 (6.9)	48.1 (45.3–51.0)	99.5 (99.4–99.6)	88.5 (85.8–90.8)	96.3 (96.0–96.5)
Black	2600	478 (18.4)	50.0 (45.5–54.5)	98.8 (98.3–99.2)	90.5 (86.3–93.5)	89.8 (88.5–90.9)
Other	3175	562 (17.7)	50.7 (46.6–54.8)	99.1 (98.7–99.4)	92.5 (89.0–95.0)	90.3 (89.2–91.4)
Ethnicity						
Hispanic	2195	492 (22.4)	50.8 (46.4–55.2)	98.5 (97.8.–99.0)	90.9 (86.9–93.8)	87.4 (85.8–88.8)
Not Hispanic	19,410	1546 (8.0)	48.9 (46.4–51.4)	99.5 (99.4–99.6)	88.9 (86.6–90.9)	95.7 (95.4–96.0)
Other/not available	1028	172 (16.7)	47.1 (39.7–54.6)	99.8 (99.1–99.9)	97.6 (90.7–99.4)	90.4 (88.3–92.1)
Hospital type						
Tertiary care	14,277	1325 (9.3)	50.0 (47.3–52.7)	99.4 (99.2–99.5)	88.7 (86.3–90.8)	95.1 (94.7–95.5)
Not tertiary care	8356	885 (10.6)	48.0 (44.7–51.3)	99.5 (99.3–99.6)	92.0 (89.1–94.1)	94.2 (93.6–94.7)
ICU stay						
No	17,514	1404 (8.0)	47.6 (45.0–50.2)	99.4 (99.3–99.6)	88.2 (85.7–90.4)	95.6 (95.3–95.9)
Yes	5119	806 (15.8)	52.0 (48.5–55.4)	99.3 (99.0–99.5)	92.9 (90.1–94.9)	91.7 (90.9–92.5)
Vital status at discharge						
Alive	21,814	1917 (8.8)	48.5 (46.3–50.8)	99.4 (99.3–99.5)	89.0 (86.9–90.8)	95.2 (94.9–95.5)
Deceased	819	293 (35.8)	53.6 (47.8–59.2)	98.9 (97.5–99.5)	96.3 (92.0–98.3)	79.3 (76.0–82.2)

COVID-19 coronavirus disease 2019, CI confidence interval, ICU intensive care unit, ICD-10 International Classification of Disease, 10ᵗʰ revision, PPV positive predictive value, NPV negative predictive value, SARS-CoV-2 RT-PCR severe acute respiratory syndrome coronavirus 2 reverse transcription polymerase chain reaction

positivity and clinical disease, may account for lower sensitivity. Lags in coding after hospital discharge and shifts in routine testing practices among hospitalized patients may also partially explain variable sensitivity. The robust specificity of ICD-10 U07.1 coding suggests that claims-based analyses may accurately capture patients with true COVID-19 disease. Epidemiological evaluations relying solely on U07.1 coding, however, may underestimate true disease burden. We acknowledge that SARS-CoV-2 RT-PCR is an imperfect "gold standard" and has itself had variable reported performance.[6] We did not have access to corroborating information from other microbiological assessments or clinical presentations. Until higher fidelity testing is available, these data from a large
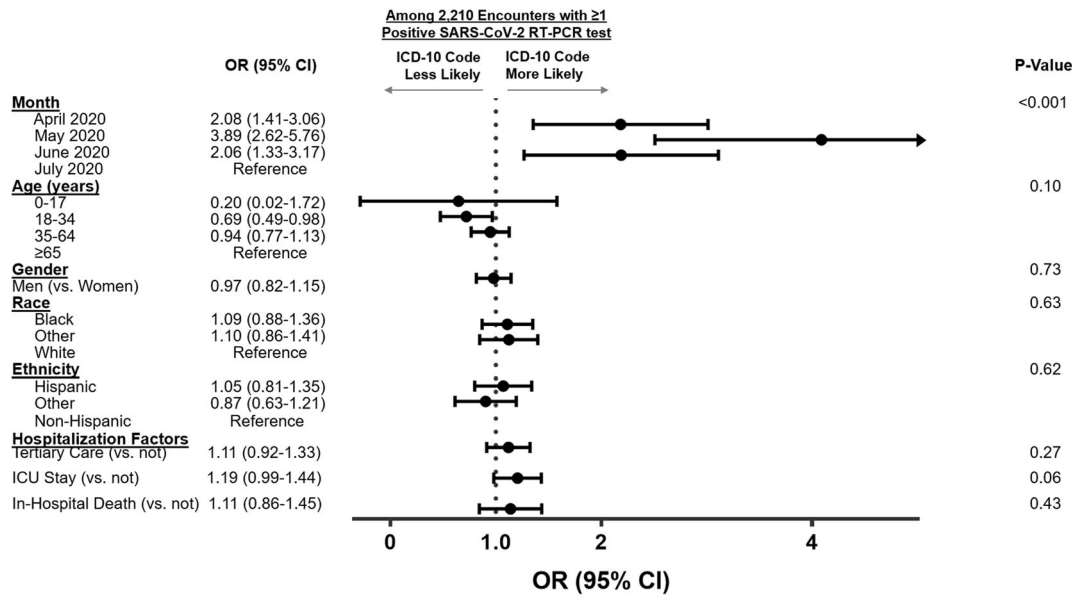
Figure 1 Independent predictors of ICD-10 sensitivity among encounters with ≥1 positive SARS-CoV-2 RT-PCR test. CI, confidence interval; ICD-10, International Classification of Disease, 10th revision; ICU, intensive care unit; OR, odds ratio; SARS-CoV-2 RT-PCR, severe acute respiratory syndrome coronavirus 2 reverse transcription polymerase chain reaction.

integrated health system inform the alignment between SARS-CoV-2 RT-PCR testing and ICD-10 coding for COVID-19.

Access to Data. The first and corresponding authors, Drs. Ankeet S. Bhatt and Muthiah Vaduganathan, had full access to the study data and take responsibility for the integrity and accuracy of its analysis.

Ankeet S. Bhatt, MD, MBA[1]
Erin E. McElrath, MBA[2]
Brian L. Claggett, PhD[1]
Deepak L. Bhatt, MD, MPH[1]
Dale S. Adler, MD[1]
Scott D. Solomon, MD[1]
Muthiah Vaduganathan, MD, MPH[1]

[1]Division of Cardiology, Department of Medicine, Brigham and Women's Hospital, Heart & Vascular Center and Harvard Medical School,
75 Francis St, Boston, MA 02115, USA
[2]Department of Medicine, Brigham and Women's Hospital,
Boston, MA, USA

Corresponding Author: Muthiah Vaduganathan, MD, MPH; Division of Cardiology, Department of Medicine, Brigham and Women's Hospital, Heart & Vascular Center and Harvard Medical School, 75 Francis St, Boston, MA 02115, USA (e-mail: mvaduganathan@bwh.harvard.edu).

Declarations:

REFERENCES

1. Centers for Disease Control and Prevention. New ICD-10-CM code for the 2019 Novel Coronavirus (COVID-19), April 1, 2020. Available at: https://www.cdc.gov/nchs/data/icd/Announcement-New-ICD-code-for-coronavirus-3-18-2020.pdf. Accessed August 15, 2020.

2. Centers for Disease Control and Prevention. ICD-10-CM Official Coding and Reporting Guidelines April 1, 2020 through September 30, 2020. Available at: https://www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf. Accessed August 15, 2020.

3. Preliminary Medicare COVID-19 Data Snapshot | CMS. Available at: https://www.cms.gov/files/document/medicare-covid-19-data-snapshot-methodology.pdf. Accessed August 15, 2020.

4. McCarthy C, Murphy S, Cohen JA, et al. Misclassification of Myocardial Injury as Myocardial Infarction: Implications for Assessing Outcomes in Value-Based Programs. JAMA Cardiol. 2019;4:460–464.

5. Crabb BT, Lyons A, Bale M, et al. Comparison of international classification of diseases and related health problems, tenth revision codes with electronic medical records among patients with symptoms of coronavirus disease 2019. JAMA Netw Open 2020;3(8):e2017703.

6. Sethuraman N, Jeremiah SS, Ryo A. Interpreting Diagnostic Tests for SARS-CoV-2. JAMA 2020;323(22):2249–2251.

Publisher's Note: *Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.*