Research Letter | Infectious Diseases

Use of and Comorbidities Associated With Diagnostic Codes for COVID-19 in US Health Insurance Claims

Kai-Cheng Yang, MS; Byungkyu Lee, PhD; Yong-Yeol Ahn, PhD; Brea L. Perry, PhD

Introduction

Accurate identification of COVID-19 diagnosis in patient medical records is essential for studies using administrative data to examine morbidity, mortality, and risk factors associated with COVID-19. Before April 1, 2020, the Centers for Disease Control and Prevention suggested using the existing International Statistical Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code B97.29 (other coronavirus as the cause of diseases classified elsewhere) as the primary diagnostic code for patients infected with COVID-19. On April 1, 2020, a new code U07.1 (2019-nCoV acute respiratory disease) was added to ICD-10-CM and was rapidly adopted by hospitals. Our study examined how nonhospital and hospital health care professionals have used these diagnostic codes in practice using a national medical claims data set in the US. We analyzed the comorbidities associated with COVID-19 diagnoses.

Figure. Trends of COVID-19 Diagnostic Codes Use From January 1 to September 30, 2020

(A) Weekly numbers of patients diagnosed with B97.29 and U07.1 codes. (B) Weekly numbers of B97.29 diagnoses from hospital and nonhospital health care professionals (individual, group practice, other facility, and unknown).

Open Access. This is an open access article distributed under the terms of the CC-BY License.
associated with COVID-19 diagnosis to assess the specificity of the legacy code and the importance of using both codes.

**Methods**

In this quality improvement study, we used the deidentified Clinformatics Data Mart Database (Optum), which comprises commercial and Medicare Advantage health plans members who are similar to the US commercially insured population with respect to demographic characteristics. Our analytic sample contained the longitudinal medical records of 28,853,694 patients across the US from January 1, 2018, to September 30, 2020. We examined the frequency of B97.29 and U07.1 in 2020 to understand their adoption by hospital and nonhospital health care professionals. We only considered the first encounter to avoid double counts. Because the legacy code may be used for other coronavirus incidents, we identified the most common co-occurring diagnoses before and after January 1, 2020, and calculated the correlation of their frequency using the SciPy package in Python, version 1.5.2. The same analysis was conducted for U07.1 for comparison. This study was approved by the Indiana University institutional review board and followed the Standards for Quality Improvement Reporting Excellence (SQUIRE) reporting guideline. Owing to the use of deidentified patient data, the need for informed consent was waived by the institutional review board.

**Results**

Of the 18,975,615 patients (mean [SD] age, 47.9 [24.1] years; 9,832,556 women [51.8%]; 9,143,059 men [48.2%]) in the data set between January 1 and September 30, 2020, 26,414 (0.14%) were

<table>
<thead>
<tr>
<th>Order</th>
<th>B97.29 in 2018-2019 (n = 4290)</th>
<th>B97.29 in 2020 (n = 63,070)</th>
<th>U07.1 in 2020 (n = 1,285,888)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I10 Essential (primary) hypertension</td>
<td>J12.89* Other viral pneumonia</td>
<td>J12.89* Other viral pneumonia</td>
</tr>
<tr>
<td>2</td>
<td>E78.5 Hyperlipidemia, unspecified</td>
<td>U07.1* 2019-nCoV acute respiratory disease</td>
<td>8982 (14.2) J96.01* Acute respiratory failure with hypoxia 219697 (17.1)</td>
</tr>
<tr>
<td>3</td>
<td>J06.9 Acute upper respiratory infection, unspecified</td>
<td>J96.01* Acute respiratory failure with hypoxia</td>
<td>8222 (13.0) I10 Essential (primary) hypertension 177265 (13.8)</td>
</tr>
<tr>
<td>4</td>
<td>J44.1 Chronic obstructive pulmonary disease with (acute) exacerbation</td>
<td>R05* Cough</td>
<td>6935 (11.0) N17.9 Acute kidney failure, unspecified 95611 (7.4)</td>
</tr>
<tr>
<td>5</td>
<td>Z87.891 Personal history of nicotine dependence</td>
<td>Z20.828* Contact with and (suspected) exposure to other viral communicable diseases</td>
<td>6633 (10.4) J18.9 Pneumonia, unspecified organism 81150 (6.3)</td>
</tr>
<tr>
<td>6</td>
<td>K21.9 Gastro-esophageal reflux disease without esophagitis</td>
<td>I10 Essential (primary) hypertension</td>
<td>6478 (10.3) E11.9 Type 2 diabetes mellitus without complications 73005 (5.7)</td>
</tr>
<tr>
<td>7</td>
<td>J96.01* Acute respiratory failure with hypoxia</td>
<td>R50.9* Fever, unspecified</td>
<td>6178 (9.8) E78.5 Hyperlipidemia, unspecified 67506 (5.2)</td>
</tr>
<tr>
<td>8</td>
<td>Z79.899 Other long term (current) drug therapy</td>
<td>J18.9 Pneumonia, unspecified organism</td>
<td>4190 (6.6) Z20.828* Contact with and (suspected) exposure to other viral communicable diseases 58513 (4.6)</td>
</tr>
<tr>
<td>9</td>
<td>I25.10 Atherosclerotic heart disease of native coronary artery without angina pectoris</td>
<td>N17.9 Acute kidney failure, unspecified</td>
<td>3679 (5.8) R05* Cough 55469 (4.3)</td>
</tr>
<tr>
<td>10</td>
<td>N17.9 Acute kidney failure, unspecified</td>
<td>E78.5 Hyperlipidemia, unspecified</td>
<td>3518 (5.6) R50.9* Fever, unspecified 47375 (3.7)</td>
</tr>
</tbody>
</table>

diagnosed with B97.29, and 279,066 (1.47%) were diagnosed with U07.1. The number of patients with a B97.29 code increased in March 2020 but rapidly diminished after the introduction of U07.1 (Figure). Although hospitals stopped using the legacy code shortly after the introduction of U07.1, some nonhospital health care professionals continued to use it (Figure). In 2020, 6 out of the 10 most frequent diagnoses that co-occurred with B97.29 were associated with COVID-19 according to the Centers for Disease Control and Prevention guideline, whereas in 2018 and 2019, only 1 out of 10 most frequent diagnoses that co-occurred with B97.29 was associated with COVID-19 (Table). The frequency of diagnostic codes that co-occurred with B97.29 in 2020 was more closely correlated with the frequency of those with U07.1 (Pearson r, 0.92; P < .001) than those with B97.29 in 2018 and 2019 (Pearson r, 0.58; P < .001). Using only U07.1 to identify patients with COVID-19 after April 1, 2020, missed 9714 patients diagnosed only with B97.29, consisting of 3.4% among 286,161 patients with either the legacy or new codes. However, the number of false positives due to screening B97.29 for patients with COVID-19 would be small, because the code’s drastic increase in 2020 can be attributed to COVID-19-related symptoms (Table).

Discussion

Using a hospital discharge data set, Kadri et al showed that the legacy code B97.29 was quickly replaced by U07.1, and its use decreased to prepandemic levels. Our quality improvement study confirmed their findings about hospitals by using large-scale medical claims data, but our results suggest that some nonhospital health care professionals have continued to use B97.29 for COVID-19 diagnoses in 2020. It is possible that patients diagnosed with B97.29 were infected with coronavirus diseases other than COVID-19, and our findings may not generalize to other data sources. However, future research on COVID-19 using claims data should consider both B97.29 and U07.1 when identifying patients with COVID-19 to avoid introducing a systematic bias across hospital and nonhospital health care professionals given the strong socioeconomic disparity in rates of COVID-19 testing and infection.

ARTICLE INFORMATION

Accepted for Publication: July 8, 2021.
Open Access: This is an open access article distributed under the terms of the CC-BY License. © 2021 Yang KC et al. JAMA Network Open.

Corresponding Author: Brea L. Perry, PhD, Department of Sociology, Indiana University-Bloomington, 1020 E Kirkwood Ave, 767 Ballantine Hall, Bloomington, IN 47405 (blperry@indiana.edu).

Author Affiliations: Luddy School of Informatics, Computing, and Engineering, Indiana University-Bloomington, Bloomington (Yang, Ahn); Department of Sociology, Indiana University-Bloomington, Bloomington (Lee, Perry).

Author Contributions: Mr Yang and Dr Perry had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Yang, Lee, Ahn.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Yang.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Yang.

Obtained funding: Ahn, Perry.

Administrative, technical, or material support: Yang, Ahn, Perry.

Supervision: Ahn, Perry.

Conflict of Interest Disclosures: None reported.
Funding/Support: This research was funded by grant ROI DA039928 from the National Institute on Drug Abuse (Dr Perry).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: The findings and conclusions in this study are those of the authors and do not necessarily represent the official position of the National Institute on Drug Abuse.

Additional Contributions: We thank Kosali Simon, PhD, and the College of Arts and Sciences at Indiana University-Bloomington for their support. They did not receive financial compensation for their contribution.

REFERENCES