ORIGINAL RESEARCH

Positive Predictive Value of COVID-19 ICD-10 Diagnosis Codes Across Calendar Time and Clinical Setting

Kristine E Lynch[1,2]
Benjamin Viernes[1,2]
Elise Gatsby[1]
Scott L DuVall [1,2]
Barbara E Jones[2,3]
Tamára L Box [4]
Craig Kreisler[4]
Makoto Jones[2,3]

[1]VA Informatics and Computing Infrastructure (VINCI), VA Salt Lake City Health Care System, Salt Lake City, UT, USA; [2]Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA; [3]Informatics, Decision-Enhancement, and Analytic Sciences (IDEAS) Center of Innovation, VA Salt Lake City Health Care System, Salt Lake City, UT, USA; [4]Analytics and Performance Integration (API), Office of Quality and Patient Safety, Veterans Health Administration, Washington, DC, USA

Purpose: To estimate the positive predictive value (PPV) of International Classification of Diseases, Tenth Revision (ICD-10) code U07.1, COVID-19 virus identified, in the Department of Veterans of Affairs (VA).

Patients and Methods: Records of ICD-10 code U07.1 from inpatient, outpatient, and emergency/urgent care settings were extracted from VA medical record data from 4/01/2020 to 3/31/2021. A weighted, random sample of 1500 records from each quarter of the one-year observation period was reviewed by study personnel to confirm active COVID-19 infection at the time of diagnosis and classify reasons for false positive records. PPV was estimated overall and compared across clinical setting and quarters.

Results: We identified 664,406 records of U07.1. Among the 1500 reviewed, 237 were false positives (PPV: 84.2%, 95% CI: 82.4–86.0). PPV ranged from 77.7% in outpatient settings to 93.8% in inpatient settings and was 83.3% in quarter 1, 80.5% in quarter 2, 86.1% in quarter 3, and 83.6% in quarter 4. The most common reasons for false positive records were history of COVID-19 (44.3%) and orders for laboratory tests (21.5%).

Conclusion: The PPV of ICD-10 code U07.1 is low, especially in outpatient settings. Directed training may improve accuracy of coding to levels that are deemed adequate for future use in surveillance efforts.

Keywords: SARS CoV-2, validation, administrative codes, electronic health records

Introduction

The need for timely, accurate, and representative healthcare data has never been more apparent since the first Coronavirus disease 2019 (COVID-19) case emerged in the US in early 2020. As one of the most commonly used nosologies, International Classification of Diseases (ICD) diagnosis codes are an appealing means for identifying and tracking cases to support healthcare surveillance efforts and facilitate epidemiologic research. They are codified, standardized, easily computable, and in some scenarios can be sufficient for understanding burden of disease. However, early in the pandemic, COVID-19 specific ICD codes did not exist. As the clinical paradigm of COVID-19 evolved throughout 2020, so too did the complexity of its clinical documentation. Due to the many sources of potential error that exist with ICD coding,[1] their utility for COVID-19 public health surveillance systems to monitor rates and trends of disease is unclear.

The first ICD-10 code for COVID-19, U07.1 [COVID-19, virus identified], was implemented in the US on April 1, 2020.[2] Further additions to COVID-19

Correspondence: Kristine E Lynch
VA Informatics and Computing Infrastructure (VINCI), VA Salt Lake City Health Care System, 500 Foothill Drive, Salt Lake City, UT, 84148, USA
Tel +1 801 582 1565 ext. 1913
Email Kristine.Lynch@va.gov

diagnostic coding were introduced in January 2021 to enable more comprehensive data capture, such as personal history of COVID-19 [Z86.6], encounter for screening for COVID-19 [Z11.52], and pneumonia due to COVID-19 [J12.82]. Although the Centers for Disease Control and Prevention (CDC) provided guidance for assigning diagnosis codes related to COVID-19 encounters,[3] the existence of guidelines does not ensure immediate, consistent, or even appropriate adoption.

Positive predictive value (PPV), the proportion of cases identified that are true cases, is one statistic used to evaluate degree of misclassification and is a commonly prioritized attribute of surveillance systems.[4] For surveillance systems that require review or investigation of identified cases, suboptimal PPV will necessitate unnecessary allocation of resources. Additionally, compromised PPV may flood the perceived case pool with non-cases making statistics such as mortality rates appear more favorable than reality. In the midst of a pandemic, where time and resources can be scarce, a surveillance system that is precise while being concurrently sufficiently sensitive is not only optimal but essential. To our knowledge, only one US study has examined PPV of code U07.1.[5] In this study, Kadri et al evaluated 52,000 hospitalizations occurring early in the pandemic from April 1, 2020 to May 31, 2020 and found the PPV of discharge diagnoses of code U07.1 to be 91.52%. Unlike sensitivity and specificity which assess the intrinsic accuracy of an instrument, PPV is population specific. It is therefore unknown whether the performance of diagnostic coding for identifying COVID-19 infection is similar for patients receiving ambulatory care, in other healthcare systems, or if it has remained stable since the code's introduction in April 2020.

The authoritative source for COVID-19 confirmed positive cases within the US Department of Veterans Affairs (VA) is the National Surveillance Tool (NST).[6,7] It was developed early in the pandemic as a way to provide VA leadership with detailed insight into cases of COVID-19 across VA medical centers in as close to real-time as possible. The NST also feeds a public facing portal that includes statistics on cases that are tested or treated in VA facilities as well as being the data provenance for cases in the VA COVID-19 Shared Data Resource, a curated data repository for VA researchers. Cases are included in the NST data feed if they have record of a positive SARS-CoV-2 PCR laboratory test in VA data or have evidence of being diagnosed with COVID-19

outside the VA healthcare system within clinical notes, which is identified and extracted via a natural language processing (NLP) system. The development and maintenance of the NLP system is resource intensive. Alternatively, augmenting VA PCR positive cases with cases identified from ICD-10 codes is an attractive approach for VA surveillance as, similar to the NLP approach, this would increase patient identification (ie, improve sensitivity) but would bypass the need to periodically update and validate the NLP system to adapt to evolving documentation patterns and clinical note templates and reduce manual review to confirm cases. However, an early informal VA assessment made clear that ICD-10 codes to identify non-VA tested or diagnosed COVID-19 infection were not sufficient for surveillance use because they were often used incorrectly. Yet, they could be integrated into future extraction processes and used as a supplement if empirical evidence demonstrated significant improvement. Whether and to what extent the performance of COVID-19 coding practices improved throughout the pandemic is unknown.

The purpose of this study was to determine the PPV of ICD-10 code U07.1 for identifying COVID-19 disease among patients at the VA. Given the likelihood of coding errors when the code was newly released due to unfamiliarity with the code, we hypothesized that PPV may have improved across time and the reasons for inaccuracies (ie, false positives) paralleled the changing clinical environments.

Materials and Methods

This study was approved in accordance with the University of Utah Institutional Review Board and with the Declaration of Helsinki. As the study was retrospective and posed no more than minimal risk to participants the requirement for written informed consent was waived. This evaluation was performed using existing data from the VA Corporate Data Warehouse (CDW), a data repository of underlying VA medical record data. It is updated nightly and receives data from over 1500 points of care across the US.[8] We performed a retrospective analysis of inpatient, outpatient, and emergency/urgent care records of ICD-10 code U07.1 occurring in VA between April 1, 2020 and March 31, 2021. We employed a stratified random sampling design to select 1500 instances of diagnosis codes to match the relative frequency of records across quarters of the one-year observation period. The four quarters were April 1, 2020–June 30, 2020, July 1,

2020–September 31, 2020, October 1, 2020–December 31, 2020, and January 1, 2021–March 31, 2021.

Using a chart abstraction tool designed specifically for this study, three trained research annotators independently reviewed 500 different instances of U07.1 (1500 in total). Annotators had access to structured data elements in the chart abstraction tool related to each patient and COVID-19 diagnosis including a summary of all VA SARS-CoV-2 lab reports (both positive and negative results), date of the first U07.1, the total number of U07.1 diagnoses, and the date vaccinated against COVID-19, if available. Additionally, annotators reviewed the clinical notes from the 30 days before and 30 days after the U07.1 record date. The tool prompted annotators to choose from seven pre-specified diagnosis reasons to classify the U07.1 record. For the purposes of this study true positives instances were categorized as 'active instances' and false positive instances were categorized as either "history of disease", "test ordered/results pending", "screening", "negative lab", "negative follow-up", or "vaccine related". Descriptions of each category are found in Table 1. The unit of observation was at the ICD-10 diagnosis instance level not the patient level. In other words, each patient could have had intervals of time throughout the one-year period with both positive and negative COVID statuses, however only the time period relevant to the diagnosis code under investigation was considered. Annotators followed CDC reporting guidelines for code U07.1 to determine case status.[3] Specifically, the CDC required healthcare systems to code:

> only a confirmed diagnosis of COVID-19 as documented by the provider, documentation of a positive COVID-19 test result, or a presumptive positive COVID-19 test result. For a confirmed diagnosis, assign code U07.1. In this context, 'confirmation' does not require documentation of the type of test performed; the provider's documentation that the individual has COVID-19 is sufficient. If the provider documents suspected, possible, probable, or inconclusive COVID-19, do not assign code U07.01.

Each of the 1500 instances was assigned to one of the seven mutually exclusive categories. A single annotator double annotated 10% of the total instances (n = 150), which included annotations only from the other two annotators. Consistency of the overall categorization (active instance versus false positive instance) was evaluated through inter-annotator agreement.

Positive predictive values were calculated overall, by quarter, and by clinical setting (inpatient, outpatient, or

Table 1 Definition of the Seven Mutually Exclusive Categories for Annotation of ICD-10 Code U07.1

Category	Definition
True Positive 1.Active COVID-19 Case	Laboratory or clinically confirmed disease, enrolled in VA telehealth COVID-19 follow-up program, or confirmation of positive COVID-19 laboratory test outside of VA.
False Positive 2.History of COVID-19 disease	Patient had COVID-19 at some point in the past but was not an active case on the diagnosis code date. The clinical text explicitly referred to patient's prior history of disease.
3.Test ordered/ results pending	COVID-19 laboratory test was ordered or patient received testing on the diagnosis code date, but the laboratory test result was not determined by that specific date.
4.COVID-19 screening	Screening information related to COVID-19 exposure and symptoms recorded in clinical note with no indication of COVID-19 case status.
5.Negative Lab	COVID-19 laboratory test was confirmed negative or the patient was informed of their negative status on the diagnosis code date.
6.Negative patient follow-up	Patient confirmed negative for COVID-19 but clinicians conducted follow-up to evaluate symptoms.
7.Vaccine related	Patient received a COVID-19 vaccination on the diagnosis code date, inquired about whether they should receive the vaccine, or reported receiving vaccination at an outside VA location.

emergency/urgent care). Confidence intervals (95% CI) were calculated using bootstrap resampling with 1000 replications.

Results

Recorded instances of diagnosis code U07.1 in VA inpatient, outpatient, and emergency/urgent care settings occurred throughout the one-year observation period with the majority being toward the end of 2020 (Figure 1). Of the total 664,406 instances, 10.0% occurred from April 1, 2020–June 30, 2020, 14.3% from July 1, 2020–September 31, 2020, 42.7% from October 1, 2020–December 31, 2020, 33.0% from January 1, 2021–March 31, 2021. This distribution led to 150, 215, 640, and 495 distinct instances manually reviewed (1500 in total) across the quarters, respectively. Although the majority of the validation sample
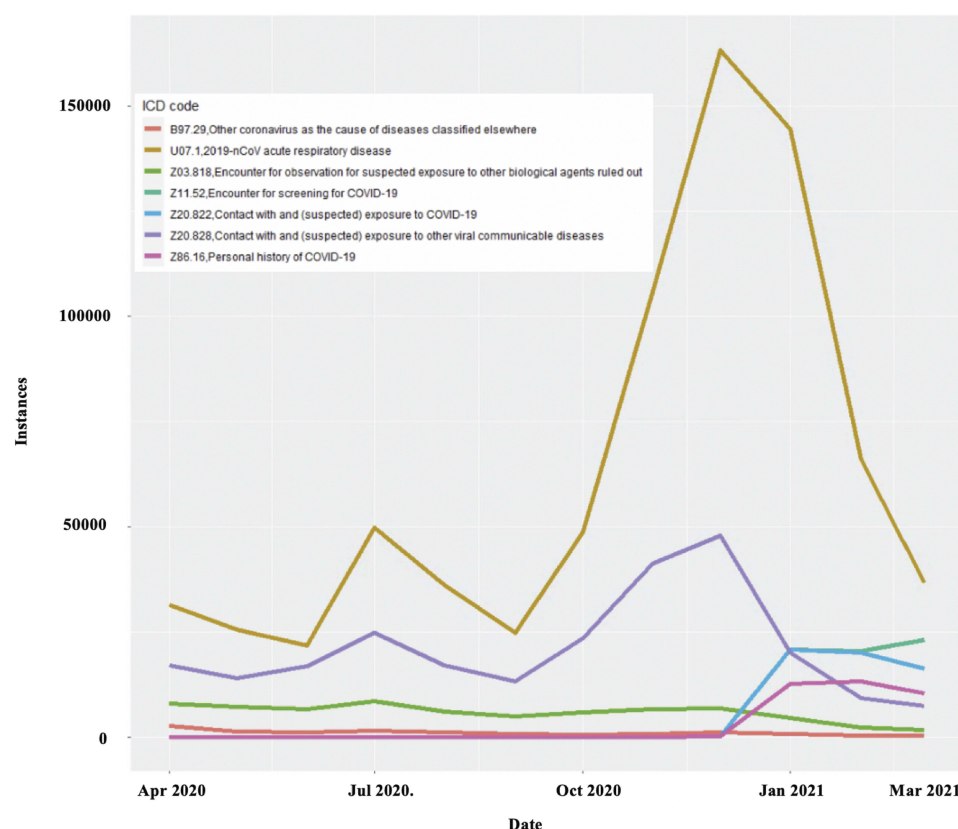
Figure 1 Distribution of instances COVID-19 related ICD-10 diagnosis codes within the Department of Veterans Affairs, April 1, 2020–March 31, 2021.

consisted of diagnoses recorded in outpatient settings (53%) followed by inpatient settings (39%), the majority of inpatient diagnoses (44%) occurred in quarter 3.

Inter-annotator agreement (IAA) was 90.7%. All disagreement between annotators was due to ambiguity in active case status documented in clinical notes. That is, annotators agreed that the patient had COVID-19 at some point in time but disagreed whether the particular instance of U07.1 covered the interval of time while the patient was active or after they had recovered. The PPV of the IAA sample was 82.7% under the original annotation and 85.3% by the second annotator. The PPV (95% CI) from April 2020–March 2021 was 84.2% (82.4–86.0) with 237 of the 1500 sampled records being false positives. The PPV was 83.3% (77.2–89.4) in the first quarter, 80.5% (75.3–85.5) in the second, 86.1% (83.4–88.8) in the third, and 83.6% (80.4–86.9) in the fourth quarter. The PPV also varied by clinical location with 81.5% (74.8–88.2) in emergency settings, 77.7% (74.9–80.5) in outpatient settings, and 93.8% (91.8–95.6) in inpatient settings (Table 2).

For the 237 false positive documentations, the most common reason throughout both the full one-year period

and in each quarter was history of COVID-19 disease, ranging from 35% of all false positives in quarter 2 to 58% in quarter 4. Documentation for an ordered test or pending laboratory test result comprised approximately 30% of the false positives from quarter 1 through quarter 3 and sharply decreased to 6% in quarter 4. Correspondingly, COVID-19 vaccination inquiries and administrations accounted for 12% of all false positives in quarter 4 whereas vaccination, which was not available until December 2020, was not a factor in the three prior quarters. A complete distribution of false positive documentation reasons is found in Table 3.

Discussion

Using manual chart review as the gold standard, we assessed the PPV of ICD-10 code U07.1 to identify patients with active COVID-19 disease across multiple clinical settings within VA from April 1, 2020 through March 31, 2021. Counter to our original hypothesis, the PPV did not improve monotonically throughout the one-year observation period, with the lowest PPV (80%) occurring in quarter 2, July–September of 2020, and the highest PPV (86%) occurring in quarter 3, October–December of 2020. Inpatient settings

Table 2 Positive Predictive Value of COVID-19 ICD-10 Code U07.1 by Quarter and Clinical Setting, Department of Veterans Affairs, April 1, 2020–March 31, 2021

	N (%)	Positive Predictive Value	95% Confidence Interval
All Time and All Settings			
April 2020–March 2021	1500 (100)	84.2	82.4–86.0
By Quarter			
April–June 2020	150 (10.0)	83.3	77.2–89.4
July–September 2020	215 (14.3)	80.5	75.3–85.5
October–December 2020	640 (42.7)	86.1	83.4–88.8
January–March 2021	495 (33.0)	83.6	80.4–86.9
By Clinical Setting			
Emergency/Urgent Care	130 (8.7)	81.5	74.8–88.2
Outpatient	794 (52.9)	77.7	74.9–80.5
Inpatient	576 (38.4)	93.8	91.8–95.6

were the most accurate while outpatient settings yielded considerably more false positives.

Although some parts of the world have shifted to a less acute phase of the pandemic, the present findings nevertheless have contemporary relevance. Vaccines are still being administered in the US and other parts of the globe, and many areas are still experiencing elevated case counts and mortality. Even among areas where vaccination is available, there are emerging variants of interest and continued cases in several regions and among specific patient demographic profiles. This suggests that surveillance efforts will continue to be warranted. However, the scope and objectives of the surveillance will dictate the minimum acceptable level of code performance. Kadri et al,[5] for example, found COVID-19 inpatient diagnosis codes were relatively high quality (PPV = 91%, sensitivity = 98%) and thus suitable for surveillance of inpatient cases and research aimed at evaluating the cost associated with a COVID-19 inpatient hospital admissions.

Using medical record data from Denmark, Bodilsen et al found DRG codes for COVID-19 (DB342A and DB972A) had exceptionally high PPV (99%) from February 2020 through May 2020 and concluded these codes to be valid for registry-based prognosis research.[9]

In situations where surveillance efforts have access to laboratory data, cases identified on the basis of positive laboratory evidence can be supplemented with the inclusion of cases identified solely by administrative claims. Although not explicitly assessed in this study, while this approach may improve sensitivity, it may do so at a detriment to the PPV. Due to the unacceptable overall false positive rate found by the present study, using ICD-10 codes, either alone or to supplement laboratory defined cases, for COVID-19 surveillance or research is ill-advised in VA. Aside from standard surveillance, ICD codes with PPV this low may also be insufficient for patient outreach efforts, where patients might be unnecessarily bothered by

Table 3 Frequency of False Positive Records of COVID-19 ICD-10 Code U07.1 by False Positive Category Across Time, Department of Veterans Affairs, April 1 2020–March 31 2021

	Time Period				
	Overall	Quarter 1	Quarter 2	Quarter 3	Quarter 4
False positive category	April 2020–March 2021	April–June 2020	July–September 2020	October–December 2020	January–March 2021
All false positives	237	25	42	89	81
History of COVID-19 disease	105 (44.3)	9 (36.0)	15 (35.7)	34 (38.2)	47 (58.0)
Test ordered/results pending	51 (21.5)	9 (36.0)	13 (30.9)	24 (27.0)	5 (6.1)
COVID-19 screening	29 (12.2)	3 (12.0)	6 (14.3)	12 (13.5)	8 (9.9)
Negative lab	34 (14.3)	3 (12.0)	7 (16.7)	13 (14.6)	11 (13.6)
Negative patient follow-up	8 (3.4)	1 (4.0)	1 (2.4)	6 (6.7)	0 (0.0)
Vaccine related	10 (4.2)	0 (0.0)	0 (0.0)	0 (0.0)	10 (12.4)

phone calls, constructing measures of re-infection, or identifying vaccine breakthrough cases.

The reasonably high PPV for inpatient COVID-19 diagnoses in the present study (94%) is similar to prior findings by Kadri and colleagues,[5] who reported a PPV of 91%. The slight difference in PPV is likely attributable to the differing observation periods whereas the current study included data from the entire first year of the pandemic while the prior study included only the first five months of 2020. Further, it is also important to note that predictive values vary as a function of disease prevalence whereas the higher the prevalence the higher the PPV, and reciprocally, the lower the prevalence the lower the PPV. Like most of the US, infection rates in VA as well as hospitalizations due to COVID-19 were highest in the latter part of 2020, which was likely a driving factor to the slightly higher PPV in quarter 3.

The modest overall PPV (84%) was influenced by the proportion of diagnoses occurring in outpatient settings. Common practice in medical record phenotyping is to define conditions according to at least one inpatient or at least two outpatient diagnoses.[10,11] This more restrictive definition increases the likelihood that individuals identified through rule-out diagnoses or data entry errors are excluded from analyses. As such, suboptimal outpatient documentation practice for COVID-19 diagnosis codes is not all that surprising despite the unique pandemic situation. Poor predictive value, however, may not apply to all COVID-19 related codes. Early pandemic research by Crabb and colleagues found relatively high predictive values despite extremely poor sensitivity for ICD-10 codes for symptoms which ranged from 0.43 for cough to 0.21 for dyspnea.[12]

We found that the introduction of additional ICD-10 codes to document a patient's COVID-19 experience did not coincide with increased precision of code U07.1 as might have been expected. For example, ICD-10 code Z86.16 [personal history of COVID-19] became effective on January 1, 2021, and although there was some degree of immediate uptake of this code in VA (Figure 1) the most common reason for false positive documentation (48%) from January 2021 through March 2021 was using U07.1 to note a patient's history of having COVID-19. Having coders and/or physicians de-adopt a practice and change their routine workflow is not a trivial task, and since the VA does not inherently rely on administrative coding for billing purposes, changes in coding practices may take even longer in VA environments.

The results presented in the study have multiple future uses. First, evaluating the PPV of administrative codes can be used to quantify the uncertainty of estimates in epidemiologic research.[13] Second, understanding the context in which coding errors occur can inform efforts to improve future documentation practices and increase the usefulness of the codes for both research and surveillance. One proposed solution for improvement in coding and documentation is education followed by audit and feedback during a code's initial roll-out.

This study has limitations that warrant discussion. First, findings may not be generalizable to other healthcare systems. The VA is the largest integrated healthcare system in the US, serving more than six million Veterans annually nationwide. The VA differs from private sector healthcare systems in multiple ways including how services are reimbursed. For instance, it runs on a capitation-based budgeting system (Veterans Equitable Resource Allocation (VERA)), which enables the prioritization of quality, access, and appropriateness of services provided over volume of billable services. As such, the incentives for coding accuracies may differ from other healthcare systems and may require a tailored approach for improvement. However, accurate uptake of ICD-10 code U07.1 likely lagged to some degree in all healthcare systems due to the off-cycle release of the ICD code, the new COVID-19 clinical environment, and strain on clinical and administrative resources. Second, sensitivity was not assessed in this study. Sensitivity could be a useful measure of validity for researchers using claims-based data or the like (when laboratory test results are unavailable) to better understand the extent of data capture/missingness when using a phenotype defined by ICD codes alone. However, a sensitivity calculation requires an enumerated denominator of true positives over time which cannot be easily directly assessed without great expense or in institutional settings.

Conclusion

The availability of COVID-19 testing expansions outside of patients' usual places of care is a critical reason that non-laboratory-based mechanisms for surveilling patients has been necessary in VA. When a patient tests positive or is diagnosed outside of the VA but seeks care within VA, structured lab data may not reach the VA medical record. These patients would not be identified in VA if a case definition included only VA lab positive patients. Supplementing laboratory data may be particularly

important for specific patient populations, such as low-income and/or rural patients, if they more heavily rely on testing sites outside the VA.

While the VA's current hybrid NLP and structured data (ie, laboratory data) approach is adequate, the use of ICD-10 codes would be ideal because of their ease of integration in a surveillance system. However, in this nationwide US study, we found ICD-10 diagnosis code U07.1 has low PPV, especially in outpatient settings, making it not sufficiently accurate for comprehensive COVID-19 surveillance. Future work should focus on interventions to improve coding practices and to standardize adoption so ICD-10 codes can be a viable option for future pandemic surveillance.

Abbreviations

COVID-19, Coronavirus disease 2019; ICD, International Classification of Diseases; CDC, Centers for Disease Control and Prevention; PPV, positive predictive value; VA, Department of Veterans Affairs; NST, National Surveillance Tool; NLP, natural language processing; CDW, corporate data warehouse; IAA, inter-annotator agreement; CI, confidence interval.

Acknowledgments

Disclosure

Dr Scott L DuVall reports grants from Anolinx, LLC, Astellas Pharma, Inc, AstraZeneca Pharmaceuticals LP, Boehringer Ingelheim International GmbH, Celgene Corporation, Eli Lilly and Company, Genentech Inc., Genomic Health, Inc., Gilead Sciences Inc., GlaxoSmithKline PLC, Innocrin Pharmaceuticals Inc., Janssen Pharmaceuticals, Inc., Kantar Health, Myriad Genetic Laboratories, Inc., Novartis International AG, and Parexel International Corporation, outside the submitted work. The author reports no other conflicts of interest in this work.

References

1. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health Serv Res*. 2005;40(5 Pt 2):1620–1639. doi:10.1111/j.1475-6773.2005.00444.x

2. Centers for Disease Control and Prevention. New ICD-10-CM for the 2019 Novel Coronavirus (COVID-19), December 3; 2020. Available from: https://www.cdc.gov/nchs/data/icd/Announcement-New-ICD-code-for-coronavirus-19-508.pdf. Accessed June 11, 2021.

3. Centers for Disease Control and Prevention. ICD-10-CM official coding and reporting guidelines: April 1, 2020 through September 30; 2020. Available from: https://www.cdc.gov/nchs/icd/icd10cm.htm. Accessed October 9, 2020.

4. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. *MMWR Recomm Rep*. 2001;50(Rr–13):1–35; quiz CE31–37.

5. Kadri SS, Gundrum J, Warner S, et al. Uptake and accuracy of the diagnosis code for COVID-19 among US hospitalizations. *JAMA Netw Open*. 2020;324(24):2553–2554.

6. Stratford D. National Surveillance Tool assesses readiness across VA's health system. VAntage Point. 2020.

7. Chapman AB, Peterson KS, Turano A, Box TL, Wallace KS, Jones M. *A natural language processing system for national COVID-19 surveillance in the US Department of Veterans Affairs*. ACL 2020 Workshop NLP-COVID Submission; 2020.

8. Fihn SD, Francis J, Clancy C, et al. Insights from advanced analytics at the Veterans Health Administration. *Health Affairs*. 2014;33 (7):1203–1211. doi:10.1377/hlthaff.2014.0054

9. Bodilsen J, Leth S, Nielsen SL, Holler JG, Benfield T, Omland LH. Positive Predictive Value of ICD-10 Diagnosis Codes for COVID-19. *Clin Epidemiol*. 2021;13:367–372. doi:10.2147/CLEP.S309840

10. Abrams TE, Vaughan-Sarrazin M, Keane TM, Richardson K. Validating administrative records in post-traumatic stress disorder. *Int J Methods Psychiatr Res*. 2016;25(1):22–32. doi:10.1002/mpr.1470

11. Nickel KB, Wallace AE, Warren DK, et al. Modification of claims-based measures improves identification of comorbidities in non-elderly women undergoing mastectomy for breast cancer: a retrospective cohort study. *BMC Health Serv Res*. 2016;16(a):388. doi:10.1186/s12913-016-1636-7

12. 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–e2017703. doi:10.1001/jamanetworkopen.2020.17703

13. Newcomer SR, Xu S, Kulldorff M, Daley MF, Fireman B, Glanz JM. A primer on quantitative bias analysis with positive predictive values in research using electronic health data. *J Am Med Inform Assoc*. 2019;26(12):1664–1674. doi:10.1093/jamia/ocz094