



RESEARCH ARTICLE

Comparing reliability of ICD-10-based COVID-19 comorbidity data to manual chart review, a retrospective cross-sectional study

Joseph W. Schaefer¹ | Joshua M. Riley¹  | Michael Li² |
Dianna R. Cheney-Peters³ | Chantel M. Venkataraman⁴ | Chris J. Li¹ |
Christa M. Smaltz⁴ | Conor G. Bradley¹ | Crystal Y. Lee¹ |
Danielle M. Fitzpatrick⁴ | David B. Ney¹ | Dina S. Zaret¹ | Divya M. Chalikonda⁴ |
Joshua D. Mairose¹ | Kashyap Chauhan⁴ | Margaret V. Szot⁴ | Robert B. Jones⁴ |
Rukaiya Bashir-Hamidu⁴ | Shuji Mitsuhashi⁴ | Alan A. Kubey^{3,5} 

¹Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania, USA

²Institute of Emerging Health Professions, Center for Digital Health and Data Science, Thomas Jefferson University, Philadelphia, Pennsylvania, USA

³Division of Hospital Medicine, Department of Medicine, Thomas Jefferson University Hospital, Philadelphia, Pennsylvania, USA

⁴Department of Medicine, Internal Medicine Residency, Thomas Jefferson University Hospital, Philadelphia, Pennsylvania, USA

⁵Division of Hospital Internal Medicine, Department of Internal Medicine, Mayo Clinic, Rochester, Minnesota, USA

Correspondence

Alan A. Kubey, Division of Hospital Medicine, Thomas Jefferson University Hospital, 833 Chestnut Street, Suite 701, Philadelphia, PA 19107, USA.

Email: alan.kubey@jefferson.edu

Abstract

International Statistical Classification of Disease and Related Health Problems, 10th Revision codes (ICD-10) are used to characterize cohort comorbidities. Recent literature does not demonstrate standardized extraction methods. Objective: Compare COVID-19 cohort manual-chart-review and ICD-10-based comorbidity data; characterize the accuracy of different methods of extracting ICD-10-code-based comorbidity, including the temporal accuracy with respect to critical time points such as day of admission. Design: Retrospective cross-sectional study. Measurements: ICD-10-based-data performance characteristics relative to manual-chart-review. Results: Discharge billing diagnoses had a sensitivity of 0.82 (95% confidence interval [CI]: 0.79–0.85; comorbidity range: 0.35–0.96). The past medical history table had a sensitivity of 0.72 (95% CI: 0.69–0.76; range: 0.44–0.87). The active problem list had a sensitivity of 0.67 (95% CI: 0.63–0.71; range: 0.47–0.71). On day of admission, the active problem list had a sensitivity of 0.58 (95% CI: 0.54–0.63; range: 0.30–0.68) and past medical history table had a sensitivity of 0.48 (95% CI: 0.43–0.53; range: 0.30–0.56). Conclusions and Relevance: ICD-10-based comorbidity data performance varies depending on comorbidity, data source, and time of retrieval; there are notable opportunities for improvement. Future researchers should clearly outline comorbidity data source and validate against manual-chart-review.

KEYWORDS

comorbidity, COVID-19, electronic health records, International Classification of Diseases, research design

1 | INTRODUCTION

Manual chart review is considered the gold standard for clinical research but requires extensive time and personnel. Using International Statistical Classification of Disease and Related Health Problems, 10th Revision codes (ICD-10) in a hospital database and/or electronic health record (EHR) is an efficient way to characterize comorbidities. Many COVID-19 studies have used ICD-10 to identify comorbidities.^{1–10} Prior COVID-19 cohort studies, however, do not demonstrate standardized practice in the extraction of ICD-10-code-based comorbidity data. The purpose of this study is to compare the comorbidities identified by manual-chart-review versus ICD-10 coding in a cohort of patients hospitalized with COVID-19. The secondary aim is to characterize the accuracy of different methods of extracting ICD-10-code-based comorbidity, including the temporal accuracy with respect to critical time points such as day of admission.

2 | METHODS

2.1 | Study design, setting, and population

This retrospective cross-sectional study included all adults (age ≥ 18 years) admitted with COVID-19 to Thomas Jefferson University Hospital (TJUH) in Philadelphia from March 1st to June 6th, 2020. COVID-19 was defined as a positive SARS-CoV-2 qualitative polymerase chain reaction. We excluded patients who were transferred from another institution, pregnant, and/or incarcerated. ICD-10 codes were obtained from the Jefferson Health datamart. ICD-10 codes were sub-grouped by their source of origin in the EHR: past medical history, active problem list, and billed discharge diagnoses (i.e., summary for insurance claims). For the Jefferson Health system, the discharge billing diagnoses are entered after an encounter by nonclinical staff; the past medical history table and active problem list are populated by clinicians. The data used for analysis was transferred to an enterprise server on September 22, 2020 (108–205 days after admission). TJUH institutional review board approved this study (IRB#: 21E.265). This study follows the reporting guidelines outlined in Strengthening the Reporting of Observational Studies in Epidemiology.¹¹

2.2 | Chart review data collection

Two independent reviewers extracted comorbidity information via manual EHR chart review. Comorbidities were included that were listed in emergency department notes, admission history and physical note, immediate past discharge summary (if existed) and/or the problem list. Comorbidities of interest were chosen through an April, 2020 literature review of COVID-19 mortality independent risk factors. For data validation, when two reviewers disagreed, a third independent reviewer adjudicated discrepancies.

2.3 | ICD-10 classification process

The “icd” package in R (Version 3.3, Author: Jack O. Wasey, MD, Children's Hospital of Pennsylvania) was used to assign comorbidities using literature-supported ICD-10 mappings established by Quan et al.^{12,13} The comorbidities analyzed for ICD-10 comparison were congestive heart failure (CHF), cerebral vascular disease (CEVD), diabetes, chronic kidney disease (CKD), cancer, human immunodeficiency virus (HIV), and hypertension (HTN). Composites of subtypes (e.g., complicated and uncomplicated diabetes) were used for diabetes, cancer, and hypertension to allow for comparison to manual review. These comorbidities were selected due to availability in both the manual review and the Quan et al.¹³ ICD-10 mappings.

Sensitivity and specificity analysis was performed on each ICD-10 source against the chart review results. The discharge diagnoses table was linked by hospital encounter identification numbers, whereas the active problem list and past medical history were linked by medical record numbers. The medical record number linked active problem list and past medical history table included a timestamp of when the code was recorded. This timestamp was used to filter the ICD-10 codes: day of admission, day of discharge, and no time restriction.

2.4 | Statistical analysis

Statistical analysis was performed using the methodology detailed by Crabb et al.¹⁴ Sensitivity, specificity, positive predictive value, and negative predictive value were calculated by comparing ICD-10-derived comorbidities with the corresponding measures from manual chart review; 95% confidence intervals (CI) were calculated by bootstrapping the point estimates using random resampling with replacement to create 1000 samples of the same size as the original group. The 95% CIs for each performance characteristic were calculated from the empirical bootstrap distribution.

3 | RESULTS

A total of 426 patients were admitted to TJUH for COVID-19 from March 17th to June 6th, 2020. The average patient age was 64.4; 43.4% were female. The percentages of Black, white, Hispanic, Asian, and other patients were 54.4%, 25.6%, 10.6%, 6.8%, and 2.6%, respectively. The mortality rate was 16.7%.

Frequency of comorbidities as determined by manual review were as follows: CHF (80), CEVD (81), diabetes (162), CKD (93), cancer (72), HIV (10), and HTN (301). Table 1 displays the performance characteristics for each ICD-10 diagnosis table from the EHR. The discharge diagnoses table (108–205 days after admission) was the most sensitive individual table for all comorbidities with sensitivity of 0.82 (95% CI: 0.79–0.85), with individual comorbidity sensitivity ranging from 0.35 to 0.96. It was followed by the past medical history table with a sensitivity of 0.72 (95% CI: 0.69–0.76), with

TABLE 1 ICD-10-based data performance

	Manual chart review count	Positive predictive value	Negative predictive value	Sensitivity	Specificity
Discharge diagnoses					
All	799	-	-	0.82 (0.79, 0.85)	0.94 (0.92, 0.96)
CHF	80	0.81 (0.73, 0.89)	0.98 (0.97, 0.99)	0.92 (0.86, 0.97)	0.95 (0.93, 0.97)
CEVD	81	0.83 (0.7, 0.94)	0.87 (0.83, 0.9)	0.36 (0.26, 0.47)	0.98 (0.97, 0.99)
Diabetes	162	0.98 (0.96, 1)	0.97 (0.95, 0.99)	0.96 (0.92, 0.99)	0.99 (0.97, 1)
CKD	93	0.72 (0.64, 0.81)	0.99 (0.97, 1)	0.96 (0.91, 0.99)	0.9 (0.86, 0.93)
Cancer	72	0.96 (0.87, 1)	0.88 (0.85, 0.91)	0.35 (0.23, 0.46)	1 (0.99, 1)
HIV	10	1 (1, 1)	0.99 (0.97, 1)	0.41 (0.1, 0.73)	1 (1, 1)
HTN	301	0.96 (0.93, 0.98)	0.84 (0.77, 0.89)	0.93 (0.9, 0.95)	0.9 (0.84, 0.95)
Past medical history					
All	799	-	-	0.72 (0.69, 0.76)	0.95 (0.93, 0.97)
CHF	80	0.92 (0.84, 1)	0.91 (0.88, 0.94)	0.58 (0.48, 0.69)	0.99 (0.98, 1)
CEVD	81	0.8 (0.69, 0.89)	0.91 (0.88, 0.94)	0.58 (0.48, 0.68)	0.97 (0.94, 0.98)
Diabetes	162	0.97 (0.93, 0.99)	0.92 (0.89, 0.95)	0.87 (0.82, 0.92)	0.98 (0.96, 1)
CKD	93	0.87 (0.76, 0.96)	0.86 (0.83, 0.89)	0.44 (0.35, 0.54)	0.98 (0.96, 0.99)
Cancer	72	0.84 (0.74, 0.94)	0.92 (0.89, 0.95)	0.6 (0.49, 0.7)	0.98 (0.96, 0.99)
HIV	10	1 (1, 1)	1 (0.99, 1)	0.8 (0.46, 1)	1 (1, 1)
HTN	301	0.96 (0.93, 0.98)	0.7 (0.62, 0.77)	0.84 (0.79, 0.88)	0.91 (0.86, 0.96)
Active problem list					
All	799	-	-	0.67 (0.63, 0.71)	0.98 (0.97, 0.99)
CHF	80	0.86 (0.77, 0.93)	0.94 (0.91, 0.96)	0.71 (0.62, 0.8)	0.97 (0.96, 0.99)
CEVD	81	0.93 (0.83, 1)	0.89 (0.86, 0.92)	0.47 (0.36, 0.58)	0.99 (0.98, 1)
Diabetes	162	0.98 (0.95, 1)	0.86 (0.82, 0.9)	0.73 (0.67, 0.8)	0.99 (0.98, 1)
CKD	93	0.91 (0.85, 0.97)	0.94 (0.91, 0.96)	0.77 (0.69, 0.85)	0.98 (0.96, 0.99)
Cancer	72	0.91 (0.82, 0.98)	0.92 (0.89, 0.95)	0.58 (0.47, 0.69)	0.99 (0.98, 1)
HIV	10	1 (1, 1)	0.99 (0.98, 1)	0.59 (0.29, 0.9)	1 (1, 1)
HTN	301	0.99 (0.97, 1)	0.55 (0.49, 0.62)	0.67 (0.62, 0.73)	0.98 (0.94, 1)
Pooled active problem list and past medical history					
All	799	-	-	0.86 (0.83, 0.89)	0.95 (0.92, 0.96)
CHF	80	0.86 (0.79, 0.93)	0.97 (0.95, 0.99)	0.87 (0.8, 0.94)	0.97 (0.95, 0.98)
CEVD	81	0.81 (0.71, 0.9)	0.94 (0.91, 0.96)	0.73 (0.63, 0.82)	0.96 (0.94, 0.98)
Diabetes	162	0.96 (0.93, 0.99)	0.96 (0.93, 0.98)	0.93 (0.89, 0.97)	0.98 (0.96, 0.99)
CKD	93	0.87 (0.8, 0.94)	0.94 (0.91, 0.97)	0.78 (0.7, 0.87)	0.97 (0.95, 0.98)
Cancer	72	0.83 (0.73, 0.92)	0.95 (0.92, 0.97)	0.74 (0.63, 0.83)	0.97 (0.95, 0.99)
HIV	10	1 (1, 1)	1 (0.99, 1)	0.9 (0.67, 1)	1 (1, 1)
HTN	301	0.96 (0.93, 0.98)	0.8 (0.73, 0.86)	0.9 (0.87, 0.94)	0.9 (0.85, 0.95)

Abbreviations: CEVD, cerebral vascular disease; CHF, congestive heart failure; CKD, chronic kidney disease; HIV, human immunodeficiency virus; HTN, hypertension.

TABLE 2 Performance stratified by time

	Manual chart review count	Positive predictive value	Negative predictive value	Sensitivity	Specificity
Active problem list—day of admission					
All	799	–	–	0.58 (0.54, 0.63)	0.99 (0.98, 0.99)
CHF	80	0.87 (0.78, 0.94)	0.93 (0.9, 0.95)	0.66 (0.56, 0.76)	0.98 (0.96, 0.99)
CEVD	81	0.92 (0.81, 1)	0.88 (0.85, 0.91)	0.42 (0.31, 0.53)	0.99 (0.98, 1)
Diabetes	162	0.98 (0.95, 1)	0.82 (0.78, 0.86)	0.64 (0.57, 0.72)	0.99 (0.98, 1)
CKD	93	0.91 (0.84, 0.97)	0.92 (0.88, 0.94)	0.68 (0.58, 0.77)	0.98 (0.96, 0.99)
Cancer	72	0.9 (0.8, 0.98)	0.91 (0.88, 0.94)	0.53 (0.41, 0.63)	0.99 (0.98, 1)
HIV	10	1 (1, 1)	0.98 (0.97, 1)	0.3 (0, 0.64)	1 (1, 1)
HTN	301	0.99 (0.97, 1)	0.49 (0.43, 0.55)	0.57 (0.52, 0.63)	0.98 (0.96, 1)
Active problem list—day of discharge					
All	799			0.66 (0.61, 0.70)	0.99 (0.98, 0.99)
CHF	80	0.87 (0.79, 0.95)	0.93 (0.91, 0.96)	0.7 (0.6, 0.79)	0.98 (0.96, 0.99)
CEVD	81	0.93 (0.83, 1)	0.89 (0.86, 0.92)	0.47 (0.36, 0.58)	0.99 (0.98, 1)
Diabetes	162	0.98 (0.95, 1)	0.85 (0.81, 0.89)	0.71 (0.64, 0.78)	0.99 (0.98, 1)
CKD	93	0.91 (0.84, 0.97)	0.94 (0.91, 0.96)	0.76 (0.67, 0.85)	0.98 (0.96, 0.99)
Cancer	72	0.91 (0.81, 0.98)	0.92 (0.89, 0.94)	0.56 (0.44, 0.67)	0.99 (0.98, 1)
HIV	10	1 (1, 1)	0.99 (0.97, 1)	0.4 (0.09, 0.73)	1 (1, 1)
HTN	301	0.99 (0.98, 1)	0.55 (0.49, 0.62)	0.67 (0.62, 0.72)	0.98 (0.96, 1)
Active problem list—no time constraint					
All	799	–	–	0.67 (0.63, 0.71)	0.98 (0.97, 0.99)
CHF	80	0.86 (0.77, 0.93)	0.94 (0.91, 0.96)	0.71 (0.62, 0.8)	0.97 (0.96, 0.99)
CEVD	81	0.93 (0.83, 1)	0.89 (0.86, 0.92)	0.47 (0.36, 0.58)	0.99 (0.98, 1)
Diabetes	162	0.98 (0.95, 1)	0.86 (0.82, 0.9)	0.73 (0.67, 0.8)	0.99 (0.98, 1)
CKD	93	0.91 (0.85, 0.97)	0.94 (0.91, 0.96)	0.77 (0.69, 0.85)	0.98 (0.96, 0.99)
Cancer	72	0.91 (0.82, 0.98)	0.92 (0.89, 0.95)	0.58 (0.47, 0.69)	0.99 (0.98, 1)
HIV	10	1 (1, 1)	0.99 (0.98, 1)	0.59 (0.29, 0.9)	1 (1, 1)
HTN	301	0.99 (0.97, 1)	0.55 (0.49, 0.62)	0.67 (0.62, 0.73)	0.98 (0.94, 1)
Past medical history—day of admission					
All	799	–	–	0.48 (0.43, 0.53)	0.98 (0.97, 0.99)
CHF	80	0.92 (0.82, 1)	0.88 (0.84, 0.91)	0.41 (0.3, 0.53)	0.99 (0.98, 1)
CEVD	81	0.76 (0.62, 0.88)	0.87 (0.83, 0.9)	0.38 (0.28, 0.49)	0.97 (0.95, 0.99)
Diabetes	162	0.97 (0.93, 1)	0.79 (0.74, 0.83)	0.56 (0.49, 0.64)	0.99 (0.98, 1)
CKD	93	0.91 (0.78, 1)	0.84 (0.8, 0.87)	0.3 (0.21, 0.4)	0.99 (0.98, 1)
Cancer	72	0.9 (0.78, 1)	0.89 (0.86, 0.92)	0.4 (0.28, 0.52)	0.99 (0.98, 1)
HIV	10	1 (1, 1)	0.99 (0.98, 1)	0.5 (0.17, 0.86)	1 (1, 1)
HTN	301	0.98 (0.95, 0.99)	0.47 (0.41, 0.53)	0.55 (0.49, 0.6)	0.97 (0.94, 0.99)
Past medical history—day of discharge					
All	799	–	–	0.52 (0.47, 0.56)	0.97 (0.95, 0.98)
CHF	80	0.92 (0.82, 1)	0.88 (0.85, 0.92)	0.44 (0.32, 0.54)	0.99 (0.98, 1)

(Continues)

TABLE 2 (Continued)

	Manual chart review count	Positive predictive value	Negative predictive value	Sensitivity	Specificity
CEVD	81	0.78 (0.65, 0.89)	0.88 (0.84, 0.91)	0.43 (0.32, 0.54)	0.97 (0.95, 0.99)
Diabetes	162	0.97 (0.93, 1)	0.8 (0.76, 0.85)	0.61 (0.53, 0.68)	0.99 (0.98, 1)
CKD	93	0.88 (0.75, 0.97)	0.84 (0.8, 0.87)	0.31 (0.23, 0.41)	0.99 (0.98, 1)
Cancer	72	0.91 (0.81, 1)	0.9 (0.87, 0.93)	0.46 (0.34, 0.57)	0.99 (0.98, 1)
HIV	10	1 (1, 1)	0.99 (0.98, 1)	0.5 (0.17, 0.86)	1 (1, 1)
HTN	301	0.96 (0.93, 0.98)	0.49 (0.43, 0.56)	0.59 (0.54, 0.65)	0.94 (0.9, 0.98)
Past medical history—no time constraint					
All	799	-	-	0.72 (0.69, 0.76)	0.95 (0.93, 0.97)
CHF	80	0.92 (0.84, 1)	0.91 (0.88, 0.94)	0.58 (0.48, 0.69)	0.99 (0.98, 1)
CEVD	81	0.8 (0.69, 0.89)	0.91 (0.88, 0.94)	0.58 (0.48, 0.68)	0.97 (0.94, 0.98)
Diabetes	162	0.97 (0.93, 0.99)	0.92 (0.89, 0.95)	0.87 (0.82, 0.92)	0.98 (0.96, 1)
CKD	93	0.87 (0.76, 0.96)	0.86 (0.83, 0.89)	0.44 (0.35, 0.54)	0.98 (0.96, 0.99)
Cancer	72	0.84 (0.74, 0.94)	0.92 (0.89, 0.95)	0.6 (0.49, 0.7)	0.98 (0.96, 0.99)
HIV	10	1 (1, 1)	1 (0.99, 1)	0.8 (0.46, 1)	1 (1, 1)
HTN	301	0.96 (0.93, 0.98)	0.7 (0.62, 0.77)	0.84 (0.79, 0.88)	0.91 (0.86, 0.96)
Pooled active problem list and past medical history—day of admission					
All	799	-	-	0.74 (0.70, 0.78)	0.97 (0.95, 0.98)
CHF	80	0.86 (0.78, 0.93)	0.95 (0.93, 0.97)	0.79 (0.7, 0.87)	0.97 (0.95, 0.99)
CEVD	81	0.81 (0.7, 0.9)	0.91 (0.88, 0.94)	0.62 (0.5, 0.72)	0.97 (0.94, 0.98)
Diabetes	162	0.97 (0.94, 0.99)	0.89 (0.85, 0.92)	0.79 (0.73, 0.86)	0.98 (0.97, 1)
CKD	93	0.88 (0.8, 0.95)	0.92 (0.89, 0.95)	0.71 (0.62, 0.8)	0.97 (0.95, 0.99)
Cancer	72	0.87 (0.77, 0.95)	0.93 (0.9, 0.96)	0.64 (0.53, 0.75)	0.98 (0.96, 0.99)
HIV	10	1 (1, 1)	0.99 (0.98, 1)	0.5 (0.17, 0.86)	1 (1, 1)
HTN	301	0.97 (0.95, 0.99)	0.64 (0.57, 0.71)	0.78 (0.73, 0.83)	0.95 (0.91, 0.98)
Pooled active problem list and past medical history—day of discharge					
All	799	-	-	0.81 (0.78, 0.84)	0.96 (0.94, 0.98)
CHF	80	0.87 (0.79, 0.94)	0.97 (0.94, 0.98)	0.85 (0.77, 0.93)	0.97 (0.95, 0.99)
CEVD	81	0.82 (0.73, 0.91)	0.93 (0.9, 0.95)	0.68 (0.57, 0.78)	0.97 (0.94, 0.98)
Diabetes	162	0.97 (0.94, 0.99)	0.92 (0.89, 0.95)	0.86 (0.8, 0.91)	0.98 (0.97, 1)
CKD	93	0.87 (0.79, 0.93)	0.94 (0.91, 0.96)	0.77 (0.69, 0.86)	0.97 (0.95, 0.98)
Cancer	72	0.87 (0.78, 0.96)	0.94 (0.91, 0.96)	0.68 (0.58, 0.78)	0.98 (0.96, 0.99)
HIV	10	1 (1, 1)	0.99 (0.98, 1)	0.6 (0.25, 0.92)	1 (1, 1)
HTN	301	0.97 (0.95, 0.99)	0.74 (0.68, 0.81)	0.87 (0.83, 0.9)	0.94 (0.89, 0.98)
Pooled active problem list and past medical history—no time constraint					
All	799	-	-	0.86 (0.83, 0.89)	0.95 (0.92, 0.96)
CHF	80	0.86 (0.79, 0.93)	0.97 (0.95, 0.99)	0.87 (0.8, 0.94)	0.97 (0.95, 0.98)
CEVD	81	0.81 (0.71, 0.9)	0.94 (0.91, 0.96)	0.73 (0.63, 0.82)	0.96 (0.94, 0.98)
Diabetes	162	0.96 (0.93, 0.99)	0.96 (0.93, 0.98)	0.93 (0.89, 0.97)	0.98 (0.96, 0.99)
CKD	93	0.87 (0.8, 0.94)	0.94 (0.91, 0.97)	0.78 (0.7, 0.87)	0.97 (0.95, 0.98)

TABLE 2 (Continued)

	Manual chart review count	Positive predictive value	Negative predictive value	Sensitivity	Specificity
Cancer	72	0.83 (0.73, 0.92)	0.95 (0.92, 0.97)	0.74 (0.63, 0.83)	0.97 (0.95, 0.99)
HIV	10	1 (1, 1)	1 (0.99, 1)	0.9 (0.67, 1)	1 (1, 1)
HTN	301	0.96 (0.93, 0.98)	0.8 (0.73, 0.86)	0.9 (0.87, 0.94)	0.9 (0.85, 0.95)

Abbreviations: CEVD, cerebral vascular disease; CHF, congestive heart failure; CKD, chronic kidney disease; HIV, human immunodeficiency virus; HTN, hypertension.

individual comorbidity sensitivity ranging from 0.44 to 0.87. The active problem list had a sensitivity of 0.67 (95% CI: 0.63–0.71), with individual comorbidity sensitivity ranging from 0.47 to 0.77. A pooled active problem list and past medical history table had a sensitivity of 0.86 (95% CI: 0.83–0.89), with individual comorbidity sensitivity ranging from 0.73 to 0.93.

The discharge diagnoses table was most sensitive for CHF (0.92), diabetes (0.96), CKD (0.96), and HTN (0.93); it was least sensitive for CEVD (0.36), cancer (0.35), and HIV (0.41). The past medical history was most sensitive for diabetes (0.87), HIV (0.80), and HTN (0.84); it was least sensitive for CHF (0.58), CEVD (0.58), CKD (0.44), and cancer (0.60). The active problem list was most sensitive for CHF (0.71), diabetes (0.73), CKD (0.77), and HTN (0.67); it was least sensitive for CEVD (0.47), cancer (0.58), and HIV (0.59).

Table 2 displays the performance characteristics for the active problem list, past medical history, and a pooled table at different time points: recorded on or before day of admission, on or before day of discharge, and no time restriction. When filtering for entries recorded on or before the day of admission, the active problem list had a sensitivity of 0.58 (95% CI: 0.54–0.63); past medical history table had a sensitivity of 0.48 (95% CI: 0.43–0.53); and the pooled table had a sensitivity of 0.74 (95% CI: 0.70–0.78). When filtering for entries recorded on or before the day of discharge, the active problem list had a sensitivity of 0.66 (95% CI: 0.61–0.70); past medical history table had a sensitivity of 0.52 (95% CI: 0.47–0.56); and the pooled table had a sensitivity of 0.81 (95% CI: 0.78–0.84).

4 | DISCUSSION

Our results demonstrate significant and concerning variability in ICD-10 performance depending on comorbidity, source of ICD-10-based data, and time of data retrieval. For retrospective research, discharge billing diagnoses was the most sensitive individual table. However, when the clinical tables—active problem list and past medical history table—were pooled, they performed similarly to the discharge diagnoses table. Several of the individual comorbidity performances were insufficient. For example, for the discharge diagnoses table, a commonly used source of ICD-10 codes, stroke, cancer, and HIV had sensitivities ranging from 0.35 to 0.41. For clinical use, accuracy during the hospital stay is important. The sensitivity of the past medical history and active problem list

decreased significantly when filtering for entries recorded on or before day of admission. The sensitivity of the active problem list on admission was 0.58 (95% CI: 0.54–0.63). If the past medical history table is included, the sensitivity on admission improves to 0.74 (95% CI: 0.70–0.78). This emphasizes the importance of a thorough medical history and referencing patient notes.

In an outpatient setting, Martin et al.,¹⁵ similarly found limitations in ICD-10-derived comorbidities when comparing EHR ICD-10 data sources, with an even lower average sensitivity and specificity for the problem list, past medical history, and encounter level (discharge diagnoses): (0.41, 0.82), (0.55, 0.76), and (0.54, 0.75), respectively. Compared to our study, this lower sensitivity and specificity may be due to differences between the inpatient and outpatient setting, differing institutional practices, comorbidities analyzed, and/or COVID-19 versus non-COVID-19 populations. In a comprehensive review of 51 conditions using ICD-9-based data, Wei et al.¹⁶ found the median sensitivity was 83% with range of 3%–100%. Compared to our study, the median sensitivity is comparable to the discharge diagnoses and pooled active problem list and past medical history. This range of sensitivities demonstrates the dramatic variability in sensitivity for conditions, and we encourage institutions to verify their ICD-10-based comorbidity data.

Of note, most risk calculators are used at initial presentation yet our data shows that the accuracy of the problem list on day of admission is lacking. The temporal accuracy of the EHR, albeit not with respect to clinically relevant time points such as day of admission or discharge, has been assessed in other studies, including a study by Schulz et al.¹⁷ that calculated the lag from diagnostic criteria being met (e.g., hemoglobin A1c $\geq 6.5\%$) to structured diagnosis populating in the EHR; the mean delay for HTN, hyperlipidemia, and diabetes was 389, 198, and 166 days, respectively. Our analysis with respect to clinically relevant time points bolsters Schulz et al.'s¹⁷ data that suggest accuracy analyses should include time as a variable. For risk and/or allocation schemes, institutions should compare and validate those schemes against the timing of live data they are prospectively using at bedside, especially if they use comorbidity in native EHR risk calculators. For retrospective research, researchers should include certain structured EHR diagnoses added for a defined period after the event.

A review of literature demonstrated variability in ICD-10 retrieval practices across institutions. Given the significant variability across sources of ICD-10-based data in our data, this is a cause for

concern. The Cleveland Clinic organizes the raw EHR data, including ICD-10 codes, into concepts from the Unified Medical Language System to inform comorbidity.^{2,18} Another study used ICD-10 codes from the problem list at the beginning of the encounter while another included used ICD-10 codes from the current and prior encounters in addition to the problem list.^{1,3} We commend these teams for clearly detailing their source as this is not always clear. To bring clarity to the literature, we encourage future publications to detail the source of the ICD-10-based data and validate its accuracy against a subset manual chart review.

Our study is limited. The data is from a single-center, using a specific EHR. It is unclear if our results can be generalized. We suspect accuracy will vary based on institutional practices, including the EHR used, expectations, and/or support for documentation and billing/coding practices. This limitation, however, underscores our conclusion that other institutions/studies should manually validate their ICD-10-based data. Further, this study is limited to a single, disease-specific cohort. Although we anticipate similar trends would emerge across populations, it is unclear if these are COVID-19-specific findings. Additionally, the comorbidities analyzed were selected based on availability in previously constructed manual-chart-review-based database. We further note that TJUH has been using its current EHR for less than 5 years which may affect discrete ICD-10-based data table reliability (i.e., based on Schultz et al.,¹⁷ we anticipate data table accuracy will improve as a patient's chart ages).

The varied performance characteristics has immediate implications for COVID-19 research and resource allocation. Multivariable analyses that identify independent risk factors for COVID-19 mortality and morbidity are being used to allocate scarce resources, such as vaccines, at a public health level. Given the variability we note in ICD-10-based comorbidity identification, these studies may over-and/or-under-emphasize certain comorbidities, which may affect resource allocation. Furthermore, some health systems are using EHR-based real-time COVID-19 risk scores to inform triage, treatment allocation, and more, potentially without knowing the performance characteristics of their own EHR comorbidity data.

5 | CONCLUSION

ICD-10 codes are often used to identify comorbidities for retrospective studies and clinical risk calculators. We note significant variation in performance depending on comorbidity, source of ICD-10-based data, and time of data retrieval. We note concerning features such as a CKD sensitivity of 0.3 when using past medical history on day of admission. The data suggests that COVID-19 ICD-10-based data should be reviewed and applied carefully in the clinical and public health realms, that future researchers should clearly outline source and time of comorbidity data and validate against manual chart review, and that the research community should determine consensus in standardized extraction, analysis, and reporting of ICD-10-based data.

FUNDING INFORMATION

The research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Dianna R. Cheney-Peters, Robert B. Jones, and Alan A. Kubey: conceived the dataset and group creation. **Joseph W. Schaefer and Alan A. Kubey:** conceived study design. **Joseph W. Schaefer, Joshua M. Riley, Michael Li, Dianna R. Cheney-Peters, Chantel M. Venkataraman, Chris J. Li, Christa M. Smaltz, Conor G. Bradley, Crystal Y. Lee, Danielle M. Fitzpatrick, David B. Ney, Dina S. Zaret, Divya M. Chalikonda, Joshua D. Mairose, Kashyap Chauhan, Margaret V. Sztot, Robert B. Jones, Rukaiya Bashir-Hamidu, Shuji Mitsuhashi, and Alan A. Kubey:** performed data collection and analyses, reviewed, edited, and finalized the manuscript, providing critical feedback and changes before submission.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Joshua M. Riley  <http://orcid.org/0000-0001-6544-1606>

Alan A. Kubey  <http://orcid.org/0000-0002-8742-9381>

REFERENCES

1. Yadaw AS, Li YC, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *Lancet Digit Health*. 2020;2(10):e516-e525.
2. Jehi L, Ji X, Milinovich A, et al. Development and validation of a model for individualized prediction of hospitalization risk in 4,536 patients with COVID-19. *PLoS One*. 2020;15(8):e0237419.
3. Petrilli CM, Jones SA, Yang J, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ*. 2020;369:m1966.
4. Vaid A, Somani S, Russak AJ, et al. Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York City: model development and validation. *J Med Internet Res*. 2020;22(11):e24018.
5. Ioannou GN, Locke E, Green P, et al. Risk factors for hospitalization, mechanical ventilation, or death among 10 131 US veterans with SARS-CoV-2 infection. *JAMA Netw Open*. 2020;3(9):e2022310.
6. Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and mortality among black patients and white patients with Covid-19. *N Engl J Med*. 2020;382(26):2534-2543.
7. Anderson MR, Geleris J, Anderson DR, et al. Body mass index and risk for intubation or death in SARS-CoV-2 INFEction. *Ann Intern Med*. 2020;173(10):782-790.
8. Tartof SY, Qian L, Hong V, et al. Obesity and mortality among patients diagnosed with COVID-19: results from an integrated health care organization. *Ann Intern Med*. 2020;173(10):773-781.
9. Yehia BR, Winegar A, Fogel R, et al. Association of race with mortality among patients hospitalized with coronavirus disease 2019

- (COVID-19) at 92 US hospitals. *JAMA Netw Open*. 2020;3(8):2018039.
10. Gu T, Mack JA, Salvatore M, et al. Characteristics associated with racial/ethnic disparities in COVID-19 outcomes in an academic health care system. *JAMA Netw Open*. 2020;3(10):2025197.
 11. STROBE Statement: Available checklists. Accessed February 3, 2021. <https://www.strobe-statement.org/index.php?id=available-checklists>
 12. icd package | R Documentation. Accessed February 3, 2021. <https://www.rdocumentation.org/packages/icd/versions/3.3>
 13. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43(11):1130-1139.
 14. 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):2017703.
 15. Martin S, Wagner J, Lupulescu-Mann N, et al. Comparison of EHR-based diagnosis documentation locations to a gold standard for risk stratification in patients with multiple chronic conditions. *Appl Clin Inform*. 2017;8(3):794-809.
 16. Wei MY, Luster JE, Chan C-L, Min L. Comprehensive review of ICD-9 code accuracies to measure multimorbidity in administrative data. *BMC Health Serv Res*. 2020;20(1):489.
 17. Schulz WL, Young HP, Coppi A, et al. Temporal relationship of computed and structured diagnoses in electronic health record data. *BMC Med Inform Decis Mak*. 2021;21:21.
 18. Milinovich A, Kattan MW. Extracting and utilizing electronic health data from Epic for research. *Ann Transl Med*. 2018;6(3):5.

How to cite this article: Schaefer JW, Riley JM, Li M, et al. Comparing reliability of ICD-10-based COVID-19 comorbidity data to manual chart review, a retrospective cross-sectional study. *J Med Virol*. 2022;94:1550-1557.
[doi:10.1002/jmv.27492](https://doi.org/10.1002/jmv.27492)