Predictive Performance of the Charlson Comorbidity Index: SNOMED CT Disease Hierarchy Versus International Classification of Diseases

Stephen P Fortin

Background

Risk adjustment for disease severity and clinical prognosis is essential to obtaining valid inferences in causal inference research. The Charlson comorbidity index (CCI) was originally developed as a weighted index of observed comorbid conditions to predict 1-year mortality risk among hospitalized patients[1]. Prior research has shown the use of the SNOMED CT Disease Hierarchy by the OHDSI Health Analytics Data-to-Evidence Suite's (HADES) FeatureExtraction package leads to a higher average CCI as compared to the Quan et al. adaptation based on International Classification of Diseases (ICD) system[2,3]. The current study compares the overlap in patient capture and predictive performance of the OHDSI (SNOMED) versus Quan et al. (ICD) adaptations of the CCI.

Methods

We identified patients with an inpatient visit between 01-01-2018 to 12-31-2018 with at least 365 days of prior observation (index = first inpatient visit) in 2 administrative claims databases; specifically, IBM® MarketScan® Multi-State Medicaid Database (MDCD); and Optum® De-Identified Clinformatics Data Mart Database – Date of Death (DOD). The CCI was measured based on observed diagnosis codes occurring at or any time prior to index. For each comorbid condition included in the CCI, the overlap in patient capture between the SNOMED versus ICD vocabularies was described; and the average difference in scores attributable to each respective comorbid condition was assessed. Logistic regression was used to develop a total of 5 models for 1-year mortality based on the following sets of dependent variables: 1) age and sex; 2) CCI (ICD vocabulary); 3) CCI (SNOMED vocabulary); 4) age, sex and CCI (ICD vocabulary); and 5) age, sex and CCI (SNOMED vocabulary). The predictive performance of each vocabulary was measured using the c statistic, measured as the area under the curve of the receiver operating characteristics curve.

Results

A total of 560,065 and 1,133,400 patients meeting the study inclusion criteria were identified in MDCD and DOD, respectively. As compared to the ICD vocabulary, the SNOMED vocabulary was associated with a higher average CCI (MDCD: 3.63 vs 3.40; and DOD: 4.44 vs 4.19). As shown in Figure 1, the relative frequency of 1-year mortality stratified by CCI was similar between vocabularies. Over 5% of the study population were categorized as having the following comorbid conditions by only the ICD (and not the SNOMED) vocabulary in either database: myocardial infarction, peripheral vascular disease, chronic pulmonary disease, and mild liver disease. Conversely, over 5% of the study population were identified as having the following comorbid conditions by only the SNOMED (but not the ICD) vocabulary: chronic pulmonary disease, diabetes with chronic complications, renal disease, and malignancy, except skin neoplasms. The degree of overlap in patient capture between vocabularies in MDCD and DOD is summarized in Figures 2 and 3, respectively. A total of 24,416 (4.4%) and 145,623 (12.8%) deaths were observed within 1 year of index in MDCD and DOD, respectively. As indicated by the c statistic, similar

predictive performance was achieved by models 1, 2 and 3 (MDCD: 0.759, 0.775 and 0.780, respectively; and DOD: 0.777, 0.764 and 0.766, respectively). A statistically significant increase in the c statistic was observed with models 4 and 5 as compared to models 1, 2 and 3. That being said, similar predictive performance was achieved between models 4 versus 5 (MDCD: 0.807 vs. 0.804; and DOD: 0.809 vs. 0.807), which were based on ICD versus SNOMED vocabularies, respectively.

Conclusion

The current study found the CCI based on the SNOMED vocabulary was associated with a higher average score as compared to the ICD vocabulary. Differences in the overlap of patient capture were especially pronounced in the following comorbidities: peripheral vascular disease, chronic pulmonary disease, mild liver disease, diabetes with chronic complications, renal disease, and malignancy, except skin neoplasms. Nevertheless, the performance of the CCI derived from ICD versus SNOMED vocabularies at predicting 1-year mortality was similar suggesting both versions of the CCI are comparable measures of clinical prognosis for risk adjustment.

References

1. Charlson ME, Pompei P, Ales KL, MacKenzie CR. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases.* 1987;40:373–83.
2. Viernes B, Lynch KE, Robison B, Gatsby E, Duvall SL, Matheny ME. SNOMED CT disease hierarchies and the Charlson comorbidity index: an analysis of OHDSI methods for determining CCI. Poster presented at: OHDSI Symposium; 2020.
3. 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;43:1130–1139.

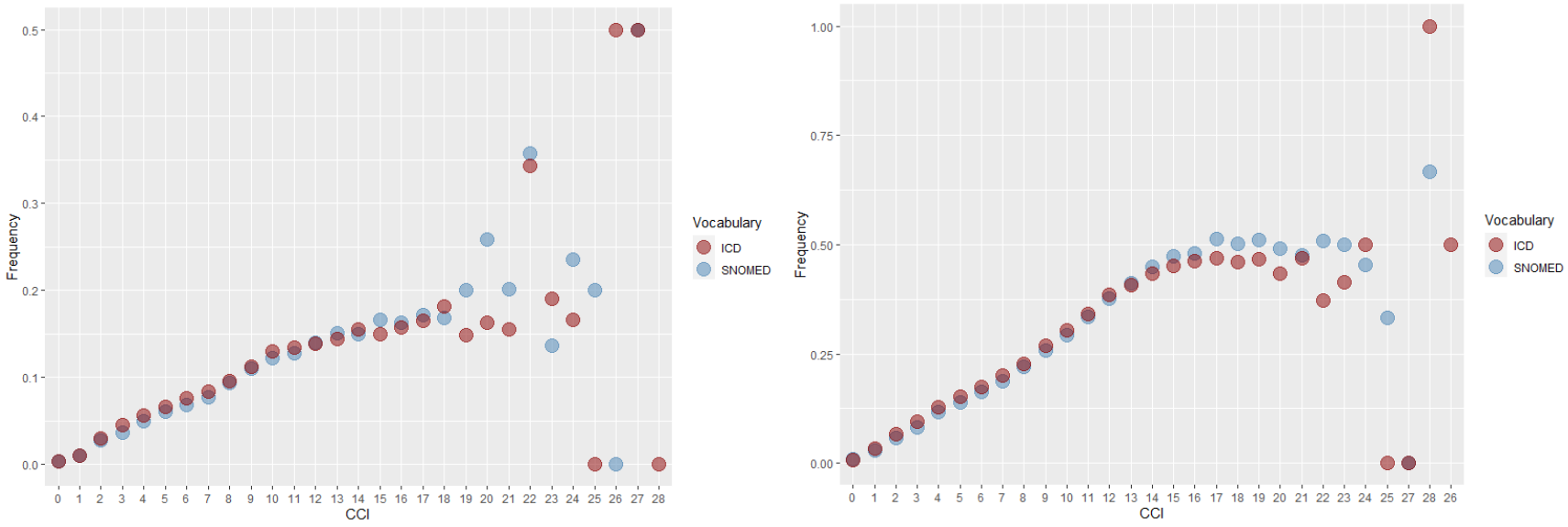
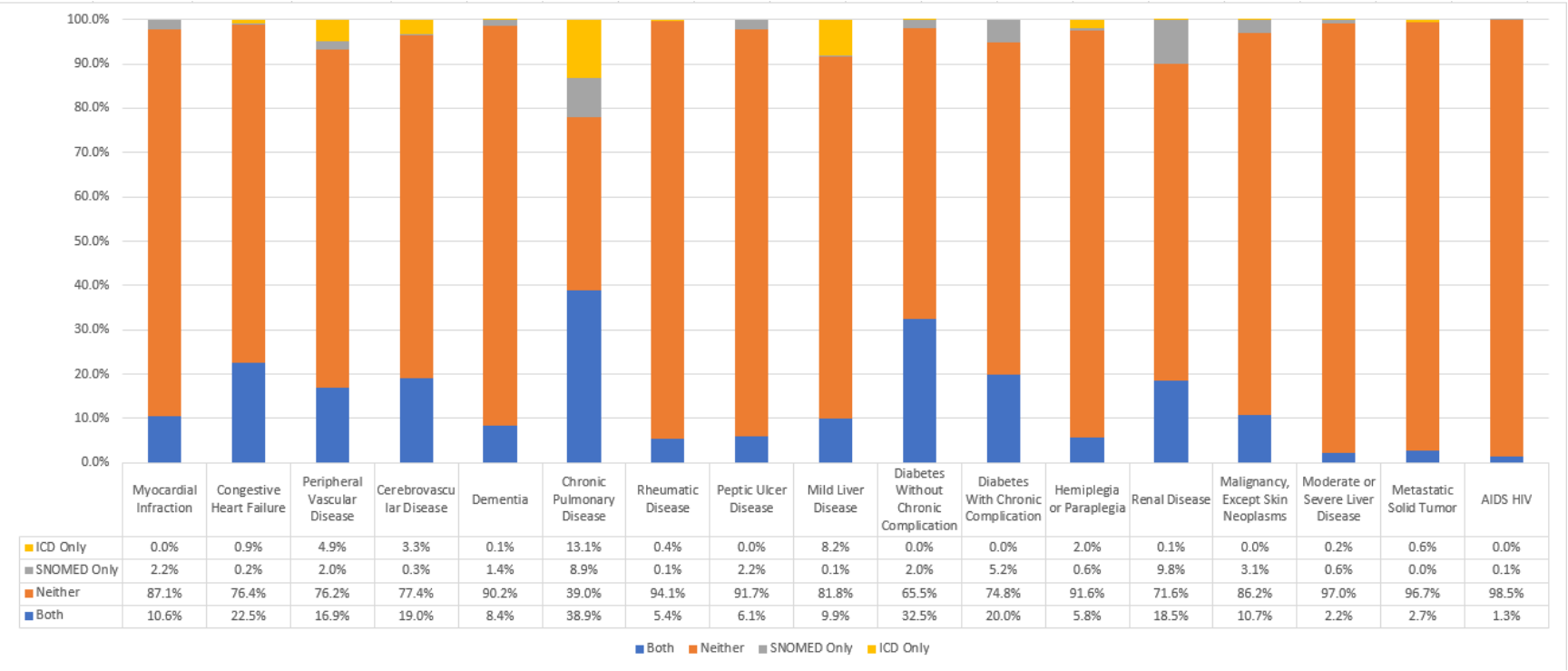
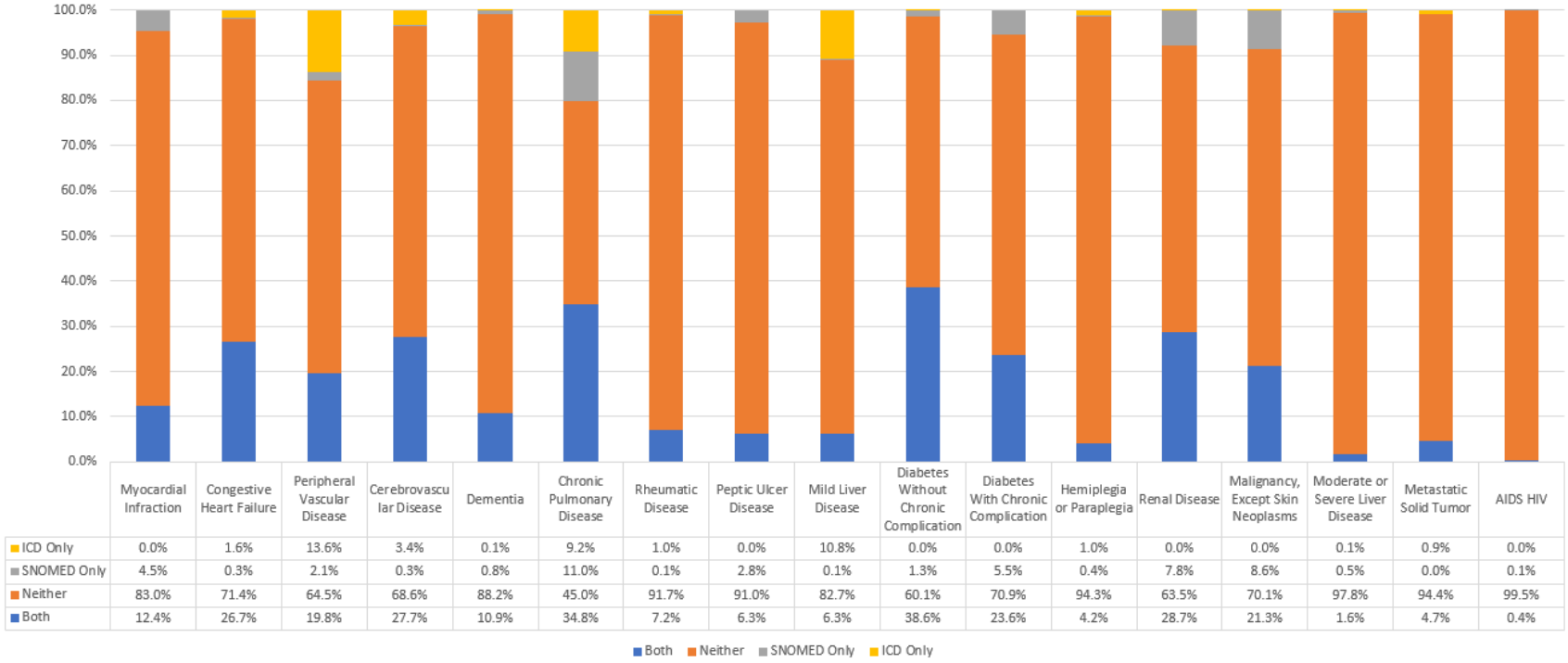


Figure 1. Relative frequency of 1-year mortality stratified by CCI in MDCD (left) and DOD (right)



	Myocardial Infraction	Congestive Heart Failure	Peripheral Vascular Disease	Cerebrovascular Disease	Dementia	Chronic Pulmonary Disease	Rheumatic Disease	Peptic Ulcer Disease	Mild Liver Disease	Diabetes Without Chronic Complication	Diabetes With Chronic Complication	Hemiplegia or Paraplegia	Renal Disease	Malignancy, Except Skin Neoplasms	Moderate or Severe Liver Disease	Metastatic Solid Tumor	AIDS HIV
ICD Only	0.0%	0.9%	4.9%	3.3%	0.1%	13.1%	0.4%	0.0%	8.2%	0.0%	0.0%	2.0%	0.1%	0.0%	0.2%	0.6%	0.0%
SNOMED Only	2.2%	0.2%	2.0%	0.3%	1.4%	8.9%	0.1%	2.2%	0.1%	2.0%	5.2%	0.6%	9.8%	3.1%	0.6%	0.0%	0.1%
Neither	87.1%	76.4%	76.2%	77.4%	90.2%	39.0%	94.1%	91.7%	81.8%	65.5%	74.8%	91.6%	71.6%	86.2%	97.0%	96.7%	98.5%
Both	10.6%	22.5%	16.9%	19.0%	8.4%	38.9%	5.4%	6.1%	9.9%	32.5%	20.0%	5.8%	18.5%	10.7%	2.2%	2.7%	1.3%

Figure 2. Degree of overlap in patient capture between SNOMED and ICD vocabularies for each comorbid condition in the CCI in MDCD



	Myocardial Infraction	Congestive Heart Failure	Peripheral Vascular Disease	Cerebrovascular Disease	Dementia	Chronic Pulmonary Disease	Rheumatic Disease	Peptic Ulcer Disease	Mild Liver Disease	Diabetes Without Chronic Complication	Diabetes With Chronic Complication	Hemiplegia or Paraplegia	Renal Disease	Malignancy, Except Skin Neoplasms	Moderate or Severe Liver Disease	Metastatic Solid Tumor	AIDS HIV
ICD Only	0.0%	1.6%	13.6%	3.4%	0.1%	9.2%	1.0%	0.0%	10.8%	0.0%	0.0%	1.0%	0.0%	0.0%	0.1%	0.9%	0.0%
SNOMED Only	4.5%	0.3%	2.1%	0.3%	0.8%	11.0%	0.1%	2.8%	0.1%	1.3%	5.5%	0.4%	7.8%	8.6%	0.5%	0.0%	0.1%
Neither	83.0%	71.4%	64.5%	68.6%	88.2%	45.0%	91.7%	91.0%	82.7%	60.1%	70.9%	94.3%	63.5%	70.1%	97.8%	94.4%	99.5%
Both	12.4%	26.7%	19.8%	27.7%	10.9%	34.8%	7.2%	6.3%	6.3%	38.6%	23.6%	4.2%	28.7%	21.3%	1.6%	4.7%	0.4%

Figure 3. Degree of overlap in patient capture between SNOMED and ICD vocabularies for each comorbid condition in the CCI in DOD