SNOMED CT Disease Hierarchies and the Charlson Comorbidity Index (CCI): An analysis of OHDSI methods for determining CCI

Benjamin Viernes, MPH[1,2], Kristine E Lynch, PhD[1,2], Brian Robison, MPH[1], Elise Gatsby, MPH[1], Scott L DuVall, PhD[1,2], Michael E Matheny, MD[3,4]

[1]VA Informatics and Computing Infrastructure, Salt Lake City, Utah, USA; [2]University of Utah, Salt Lake City, Utah, USA; [3]Vanderbilt University Medical Center, Nashville, TN, USA; [4]Tennessee Valley Healthcare System, Nashville, TN, USA

Abstract

The Charlson Comorbidity index (CCI) is one of the most widely used scores for predicting mortality and as a proxy covariate for health status in observational studies. The current quality and availability of electronic health record (EHR) data and the constantly evolving OMOP CDM has made the calculation of the CCI score for large cohorts easier than at any point since its first introduction in 1987. This study compared the OHDSI method used to calculate CCI, which uses SNOMED-CT codes, to the Quan et al (2005) method for ICD-10-CM era codes. Results demonstrate that these two methods can result in different CCI scores for the same cohort of patients. Further research is necessary to understand whether these differences cause practical issues in prediction of mortality and morbidity.

Research Category

Observational data management

Introduction

Published in 1987 by Charlson et al[1], the original paper defining the Charlson comorbidity index (CCI) has been cited almost 10,000 times. Its widespread use in research has demonstrated its utility. Further, as diagnostic coding and electronic medical records (EHRs) have become widely used in healthcare, the ability to define a patient's risk of mortality based on CCI has become simple, even for large cohorts. Most papers that have outlined methods for defining CCI from registry or EHR data, have used International Classification of Diseases (ICD) coding[2,3]. The method provided by Observational Health Data Sciences and Informatics (OHDSI) in the R package, *FeatureExtraction*[4], uses SNOMED-CT codes (and their descendants) to define conditions that qualify for inclusion in the CCI categories. This abstract will refer to this method as the OHDSI method.

The CCI contains 17 categories which contribute points towards the index score. Categories are weighted based on the probability of mortality for that category[1,2,3]. Our Department of Veterans Affairs (VA) research group, the VA Informatics and Computing Infrastructure (VINCI), found higher average CCI among a cohort of patients using the FeatureExtraction method on the VA's OMOP version 5 instance (VA OMOP) compared to Quan (2005) et al's ICD10 method (Quan method) on the VA source data of the same OMOP instance. The Quan method outlines the specific ICD10 codes that may be used to define conditions for each of the 17 CCI categories. We evaluated the magnitude and characteristics of these differences in order to inform future use and calculation of CCI within the OMOP CDM.

Methods

The VA OMOP, with health records for approximately 10 million veterans was used in conjunction with its source data, VA's Corporate Data Warehouse (CDW). 10,000 patients were randomly selected to a cohort to use for comparison between ICD10-based Quan method for CCI in CDW versus the SNOMED-based OHDSI method for CCI in VA OMOP. To be included in the analysis, patients were also required to have at least 10 conditions recorded after 10/01/2015 (when VA moved to ICD10) to ensure that the cohort would have a relatively low percentage of patients with a CCI of 0. The CCI was calculated using both methods for all patients' conditions recorded between 10/01/2015 and 01/01/2020. Records for all patients in the sample were pulled from the VA OMOP CONDITION_OCCURRENCE table. The Quan method used Condition_Source_Concept_ID to identify CCI concepts using ICD10 while the OHDSI method used the Condition_Concept_ID to identify CCI concepts

using SNOMED.

Patients with any difference in CCI score between methods were compared both by disease categories that contributed to discrepant scores, and more granularly, by the contributing ICD10 codes and SNOMED codes. In the case of a higher value from one method, the contributing methods' ICD10/SNOMED relationship in the OMOP CONCEPT_RELATIONSHIP table was evaluated.

Results

Of the 10,000 randomly sampled patients, the Quan method identified 6,231 patients with at least one CCI condition (i.e.,CCI>0) versus 6,328 patients using the OHDSI method. 403 patients were only found using the Quan method while 500 patients were only found using the OHDSI method. The mean CCIs for all patients were 3.15 and 3.47, respectively. Patients with a CCI of 7 or greater, which equates to a 0% estimated 10-year survival rate[5], also differed between groups (figure 1).

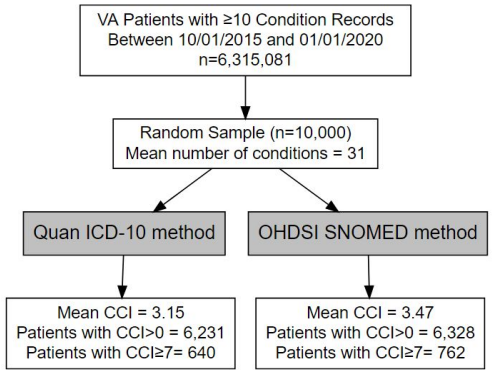

Figure 1. CCI Score differences by method

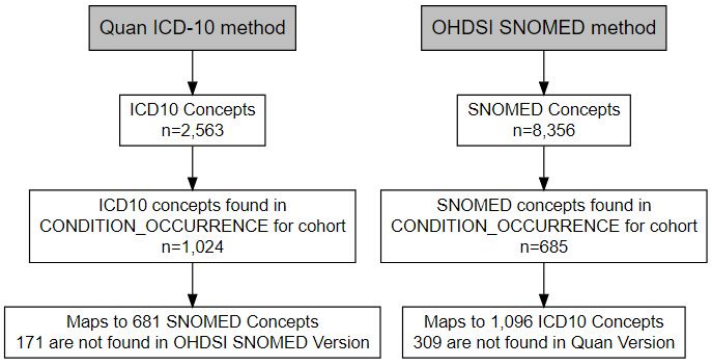

Figure 2. Mapping differences by Method

Conclusion

The differences in CCI score between methods can be attributed to the differences in ICD10 coding categories in the Quan method versus the OHDSI method of using SNOMED disease categories and their hierarchical descendants to identify CCI conditions. The 17 CCI disease categories are identified by the OHDSI method using 30 SNOMED disease categories, which equate to 8,356 SNOMED descendants. Of those that contribute CCI conditions to our cohort, they map to over 300 ICD10 codes that are not identified in the ICD10 categories identified by Quan (2005) et al. On the other side, Quan's ICD10 method identifies 2,563 unique ICD10 codes. These ICD10 codes map to 681 SNOMED concepts, of which 171 are not identified in the OHDSI SNOMED descendants. These differences and some overlap between the SNOMED disease categories in the OHDSI method (i.e.-- a patient with a metastatic solid tumor would be counted in both the 'metastatic solid tumor' and the 'any malignancy' categories due to common descendants of disease categories) are the root cause for differences between methods.

Researchers that seek to use CCI should evaluate these discrepancies to identify and validate conditions within each category using clinical expertise. More research is also necessary to understand the implications these differences have on the probabilities of mortality that CCI is generally used to predict.

References

1. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8
2. 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. doi:10.1097/01.mlr.0000182534.19832.83
3. Quan H, Li B, Couris CM, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol. 2011;173(6):676-682. doi:10.1093/aje/kwq433
4. Software. FeatureExtraction. Observational Health Data Sciences and Informatics. https://github.com/OHDSI/FeatureExtraction
5. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area [published online ahead of print, 2020 Apr 22] [published correction appears in doi: 10.1001/jama.2020.7681]. JAMA. 2020;323(20):2052-2059. doi:10.1001/jama.2020.6775