Comparing the Performance Characteristics of General and Specific Diagnosis Codes in Phenotype Algorithms.

Joel N. Swerdel, PhD MS MPH[1,3]; Patrick B. Ryan, PhD[1,2,3]

[1] Janssen Research and Development, Titusville, NJ, USA

[2] Columbia University, New York, NY, USA

[3] Observational Health Data Sciences and Informatics (OHDSI), New York, NY

Abstract - ***Background****: Phenotype algorithms (PAs) are commonly used to determine subjects with specific health conditions in observational research. PAs use sets of one or more administrative health care codes found in the data. These codes vary from general, e.g., "Acute Myocardial Infarction" to specific, e.g., "Acute Myocardial Infarction of the Anterolateral Wall". It is believed that more specific codes may reflect higher confidence of the diagnosis. The objective of this research was to use PheValuator, a package within the OHDSI toolset, to provide empirical evidence for the performance characteristics, e.g., sensitivity and positive predictive value (PPV), of PAs using both general and specific diagnosis codes compared to those using only specific codes.* ***Method:*** *Data for this study were collected between January 1, 2010 and December 31, 2018 from 4 administrative claims data sets: IBM MarketScan's Commercial Claims and Encounters, Medicare Supplemental Beneficiaries, and Multi-State Medicaid; OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN); and Optum© de-identified Electronic Health Record Dataset (OptumInsight, Eden Prairie,MN). For this research, we examined the performance characteristics of PAs for four health conditions: acute myocardial infarction, cerebral infarction, atrial fibrillation, and chronic kidney disease. We used within each PA either a code set that was a combination of several general and specific codes or a set derived from a single specific code.* ***Results:*** *We found that PAs using a single specific code consistently demonstrated higher PPVs than those using a combination of several general and specific codes. The PAs using a specific code showed lower sensitivities than those using both general and specific codes.*

Conclusion: *The results of this study provide empirical evidence that specific diagnosis codes in the patient record more likely reflect the actual disease state of the patient as measured by PPV. These results may help to improve the accuracy of PAs for defining cohorts of subjects with the health outcome of interest.*

Background: The primary approach for defining disease in observational healthcare databases is to construct phenotype algorithms (PAs), rule-based heuristics predicated on the presence, absence, and temporal logic of clinical observations. For disease states, the PAs using administrative claims datasets are based on a set of diagnosis codes used for billing purposes. The choice of codes will affect the performance of the PA. Diagnosis codes used to define a disease state may be general such as "Acute Myocardial Infarction" or specific, such as "Acute Myocardial Infarction of the Anterolateral Wall". PAs often use a set of codes that are a combination of general and specific codes. The rationale for a medical facility using a general code vs. a specific code may reflect the level of confidence in the diagnosis of the disease at a particular time in the patient treatment. Conventional belief is that general diagnoses, and, subsequently, its translation into general diagnosis codes, may be more likely provided by the clinician at the early stages of treatment possibly as a "Rule Out" code; specific diagnoses may be more likely used after the initial diagnosis has been confirmed. If that belief is correct, the positive predictive value (PPV) for a specific code should be higher than that for a general code. To date, there is little or no empirical evidence to support this belief. The objective of this research was to use PheValuator, a package within the OHDSI toolset, to provide empirical evidence for the performance characteristics, e.g., sensitivity and positive predictive value (PPV), of PAs using both general and specific diagnosis codes vs. those using only specific codes.

Methods: Data for this study were collected between January 1, 2010 and December 31, 2018 from four US administrative claims data sets: IBM MarketScan's Commercial Claims and Encounters, Medicare Supplemental Beneficiaries, and Multi-State Medicaid; and OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN).

We used two PAs in this study: 1) a PA where subjects were included if a diagnosis code appeared 1 or more times within the health record ("≥1 X Outcome") and 2) a PA where subjects were included in the cohort if a diagnosis code appeared 1 time in the patient record followed by a second diagnosis within 5 days of the first occurrence ("≥2 X Outcome"). We varied the set of codes used in each PA. For the general and specific code sets, we included the "parent" code, e.g. "Acute myocardial infarction" and all the "children" codes, e.g., "Acute Myocardial Infarction of the Anterolateral Wall". For the specific only code sets, we included only an individual "child" code. We examined the performance characteristics in four health conditions: acute myocardial infarction (1 general code set, 4 specific code sets), cerebral infarction (1 general, 4 specific), atrial fibrillation (1 general, 3 specific), and chronic kidney disease (1 general, 7 specific).

PheValuator is an R package within the OHDSI toolset that allows for evaluating the performance characteristics of PAs[1]. The tool uses diagnostic predictive modeling to determine the probability of a health condition in a large set

of subjects, the "evaluation cohort". The evaluation cohort of subjects may be used to test PAs for sensitivity, specificity, and positive and negative predictive value. We tested the PAs in this study using PheValuator with an evaluation cohort where the predicted probability for the overarching disease state, e.g. acute myocardial infarction, was determined for each subject.

Results: The performance characteristic results of the two PAs for each of the four health conditions are shown in Table 1. All values were calculated as the mean across the four datasets tested. We found that PAs using specific codes only consistently had higher PPVs than those using both general and specific codes. For example, the "≥1 X Outcome" PA using the specific diagnosis code "Acute Myocardial Infarction of the Inferolateral Wall" showed a PPV of 0.72 compared to 0.44 for the general and specific code set. The exception to this general finding was in chronic kidney disease (CKD) where the PAs using the specific codes for early stage CKD, i.e., stage 1 or 2, were similar to the PAs using the general and specific code sets. The PAs using specific only codes showed much lower sensitivities than those using both general and specific codes. For example, the "≥1 X Outcome" PA using the specific diagnosis code "Acute Myocardial Infarction of the Inferolateral Wall" showed a sensitivity of 0.02 compared to 0.67 for the general and specific code set.

Conclusion: The results of this study provide empirical evidence that specific diagnosis codes in the patient record more likely reflect the actual disease state of the patient as measured by PPV determined by PheValuator. These results may help to improve the accuracy of PAs for defining cohorts of subjects with the health outcome of interest.

Table 1: Performance Characteristics of Two Phenotype Algorithms for Four Health Conditions with Varying Diagnostic Condition Codes Sets using PheValuator

Phenotype Algorithm	Sens	PPV	Spec	NPV	Phenotype Algorithm	Sens	PPV	Spec	NPV
≥1 X MI (General+Specific)	0.665	0.441	0.983	0.992	≥ 1 X CKD (General+Specific)	0.787	0.540	0.953	0.983
≥1 X MI Inferior Wall	0.086	0.631	0.999	0.981	≥ 1 X CKD Stage 1	0.064	0.527	0.997	0.930
≥1 X MI Anterior Wall	0.070	0.613	0.999	0.980	≥ 1 X CKD Stage 2	0.184	0.556	0.990	0.937
≥1 X MI Anterolateral Wall	0.042	0.625	0.999	0.980	≥ 1 X CKD Stage 3	0.516	0.648	0.977	0.964
≥1 X MI Inferolateral Wall	0.020	0.722	0.999	0.979	≥ 1 X CKD Stage 4	0.183	0.893	0.998	0.938
≥2 X MI (General+Specific)	0.474	0.596	0.994	0.988	≥ 1 X CKD Stage 5	0.086	0.950	0.999	0.930
≥2 X MI Inferior Wall	0.040	0.810	0.999	0.980	≥ 1 X CKD Due to Hypertension	0.382	0.787	0.994	0.954
≥2 X MI Anterior Wall	0.028	0.797	0.999	0.979	≥ 1 X CKD Due to Diabetes	0.224	0.737	0.995	0.941
≥2 X MI Anterolateral Wall	0.011	0.767	0.999	0.979	≥ 2 X CKD (General+Specific)	0.720	0.622	0.968	0.978
≥2 X MI Inferolateral Wall	0.007	0.811	0.999	0.979	≥ 2 X CKD Stage 1	0.035	0.546	0.998	0.928
					≥ 2 X CKD Stage 2	0.117	0.587	0.994	0.933
≥ 1 X Cerebral Infarction (General+Specific)	0.771	0.353	0.973	0.995	≥ 2 X CKD Stage 3	0.425	0.710	0.986	0.958
≥ 1 X Cerebral infarction due to embolism of middle cerebral artery	0.053	0.771	0.999	0.982	≥ 2 X CKD Stage 4	0.140	0.932	0.999	0.935
≥ 1 X Cerebral infarction due to thrombosis of middle cerebral artery	0.034	0.743	0.999	0.981	≥ 2 X CKD Stage 5	0.060	0.977	0.999	0.929
≥ 1 X Cerebral infarct due to thrombosis of precerebral arteries	0.032	0.636	0.999	0.981	≥ 2 X CKD Due to Hypertension	0.265	0.830	0.996	0.945
≥ 1 X Cerebral infarction due to embolism of precerebral arteries	0.025	0.671	0.999	0.981	≥ 2 X CKD Due to Diabetes	0.168	0.782	0.997	0.938
≥ 2 X Cerebral Infarction (General+Specific)	0.614	0.497	0.988	0.992					
≥ 2 X Cerebral infarction due to embolism of middle cerebral artery	0.020	0.788	0.999	0.981	≥ 1 X AFib (General+Specific)	0.708	0.474	0.963	0.985
≥ 2 X Cerebral infarction due to thrombosis of middle cerebral artery	0.012	0.723	0.999	0.981	≥ 1 X Paroxysmal AFib	0.345	0.595	0.989	0.962
≥ 2 X Cerebral infarct due to thrombosis of precerebral arteries	0.012	0.651	0.999	0.981	≥ 1 X Chronic AFib	0.206	0.713	0.995	0.957
≥ 2 X Cerebral infarction due to embolism of precerebral arteries	0.009	0.688	0.999	0.981	≥ 1 X Persistent AFib	0.123	0.795	0.998	0.951
					≥ 2 X AFib (General+Specific)	0.652	0.556	0.974	0.982
					≥ 2 X Paroxysmal AFib	0.279	0.645	0.992	0.959
					≥ 2 X Chronic AFib	0.151	0.762	0.997	0.954
					≥ 2 X Persistent AFib	0.078	0.841	0.999	0.948

Sens - Sensitivity; PPV - Positive Predictive Value; Spec - Specificity; NPV - Negative Predictive Value; ≥1 X - ≥1 X Health Condition; ≥2 X - ≥2 X Health Condition; MI - Acute Myocardial Infarction; CKD - Chronic Kidney Disease; AFib - Atrial Fibrillation

Values represent the mean across 4 datasets, IBM MarketScan's Commercial Claims and Encounters, Medicare Supplemental Beneficiaries, and Multi-State Medicaid; and OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN). The continuous 3-color heat map for the data in the table was defined as Red (value = 0), Yellow (value = 0.5), and Green (value = 1).

1. PheValuator. https://githubcom/OHDSI/PheValuator 2019 [cited 2019; Available from: https://github.com/OHDSI/PheValuator