Title: #41 - Comparing the Performance Characteristics of Broad-based and Explicit Diagnosis Codes in Phenotype Algorithms.

PRESENTER: Joel Swerdel

INTRODUCTION

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 broad-based, e.g., "Acute Myocardial Infarction" to explicit, e.g., "Acute Myocardial Infarction of the Anterolateral Wall". It is believed that more explicit 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 broad-based and explicit diagnosis codes compared to those using only explicit codes.**

METHODS

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). Each dataset was converted to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), version 5.01 and has been reviewed by the New England Institutional Review Board (IRB) and were determined to be exempt from broad IRB approval. 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 broad-based and explicit codes or a set derived from a single explicit code.

RESULTS

We found that PAs using a single explicit code consistently demonstrated higher PPVs than those using a combination of several broad-based and explicit codes. The PAs using a explicit code showed lower sensitivities than those using both broad-based and explicit codes.

CONCLUSIONS

The results of this study provide empirical evidence that explicit 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.**

Cohort Definitions using **explicit diagnosis codes** have **higher positive predictive values** than definitions using broad-based codes.

Take a picture to download more information

Myocardial Infarction

Phenotype Algorithm	Sens	PPV	Spec	NPV
>=1 x MI (General and Specific)	0.665	0.441	0.983	0.992
>=1 x MI Inferior Wall	0.086	0.631	0.999	0.981
>=1 x MI Anterior Wall	0.070	0.613	0.999	0.980
>=1 x MI Anterolateral Wall	0.042	0.625	0.999	0.980
>=1 x MI Inferolateral Wall	0.020	0.722	0.999	0.979
>=2 x MI (General and Specific)	0.474	0.596	0.994	0.988
>=2 x MI Inferior Wall	0.117	0.765	0.997	0.982
>=2 x MI Anterior Wall	0.030	0.802	0.999	0.979
>=2 x MI Anterolateral Wall	0.011	0.767	0.999	0.979
>=2 x MI Inferolateral Wall	0.007	0.811	0.999	0.979

Cerebral Infarction

Phenotype Algorithm	Sens	PPV	Spec	NPV
>=1 x Cerebral Infarction (General and Specific)	0.771	0.353	0.973	0.995
>=1 x Cerebral infarction due to embolism of middle cerebral artery	0.053	0.771	0.999	0.982
>=1 x Cerebral infarction due to thrombosis of middle cerebral artery	0.034	0.743	0.999	0.981
>=1 x Cerebral infarct due to thrombosis of precerebral arteries	0.032	0.636	0.999	0.981
>=1 x Cerebral infarction due to embolism of precerebral arteries	0.025	0.671	0.999	0.981
>=2 x Cerebral Infarction (General and Specific)	0.614	0.497	0.988	0.992
>=2 x Cerebral infarction due to embolism of middle cerebral artery	0.139	0.736	0.995	0.985
>=2 x Cerebral infarction due to thrombosis of middle cerebral artery	0.014	0.727	0.999	0.981
>=2 x Cerebral infarct due to thrombosis of precerebral arteries	0.012	0.651	0.999	0.981
>=2 x Cerebral infarction due to embolism of precerebral arteries	0.009	0.688	0.999	0.981

Atrial Fibrillation

Phenotype Algorithm	Sens	PPV	Spec	NPV
>=1 X AFib (General and Specific)	0.708	0.474	0.963	0.985
>=1 X Paroxysmal AFib	0.345	0.595	0.989	0.962
>=1 X Chronic AFib	0.206	0.713	0.995	0.957
>=1 X Persistent AFib	0.123	0.795	0.998	0.951
>=2 X AFib (General and Specific)	0.652	0.556	0.974	0.982
>=2 X Paroxysmal AFib	0.377	0.632	0.984	0.971
>=2 X Chronic AFib	0.167	0.742	0.995	0.956
>=2 X Persistent AFib	0.078	0.841	0.999	0.948

Chronic Kidney Disease

Phenotype Algorithm	Sens	PPV	Spec	NPV
>=1 x CKD (General and Specific)	0.787	0.540	0.953	0.983
>=1 x CKD Stage 1	0.064	0.527	0.997	0.930
>=1 x CKD Stage 2	0.184	0.556	0.990	0.937
>=1 x CKD Stage 3	0.516	0.648	0.977	0.964
>=1 x CKD Stage 4	0.183	0.893	0.998	0.938
>=1 x CKD Stage 5	0.086	0.950	0.999	0.930
>=1 x CKD Due to Hypertension	0.452	0.749	0.983	0.961
>=1 x CKD Due to Diabetes	0.253	0.740	0.994	0.944
>=2 x CKD (General and Specific)	0.720	0.622	0.968	0.978
>=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
>=2 x CKD Stage 3	0.425	0.710	0.986	0.958
>=2 x CKD Stage 4	0.140	0.932	0.999	0.935
>=2 x CKD Stage 5	0.060	0.977	0.999	0.929
>=2 x CKD Due to Hypertension	0.265	0.830	0.996	0.945
>=2 x CKD Due to Diabetes	0.168	0.782	0.997	0.938

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