Agreement between measurement and diagnosis-based phenotype algorithms

Please list all authors and their affiliations in the correct order
Azza Shoaibi[1,2], Gowtham Rao[1,2], Dmytro Dymshyts[1,2], Anna Ostropolets[2,3], Patrick Ryan[1,2]
[1] Janssen R&D, [2] Observational Health Data Sciences and Informatics, [3]Odysseus Data Services Inc.,

Background

Various types of clinical information, including diagnoses, medications, and procedures can be used to identify a specific clinical condition or event in observational data. Previous research indicates that the accuracy of phenotyping algorithms can improve when multiple data types are incorporated[1]. However, the added benefit of incorporating clinical measurements, such as laboratory tests and their results, into such algorithms remains unclear[2]. The aim of this paper is to compare diagnosis-based phenotyping algorithms with those that are based on clinical measurements across five different clinical conditions in seven separate data sources.

Methods

We selected five condition phenotypes: rhabdomyolysis, neutropenia, thrombocytopenia, pancytopenia, and end-stage renal disease. Neutropenia, thrombocytopenia, and pancytopenia are conditions characterized by abnormal levels of blood components and are diagnosed based on measurements of such. Similarly, the diagnosis of rhabdomyolysis and end-stage renal disease rely, in part, on specific biochemical markers: creatine kinase (CK) and glomerular filtration rate (GFR) respectively.

We developed two different types of algorithms for all five conditions. Measurement-based phenotype algorithms used defined value thresholds for diagnostic markers. Diagnosis-based algorithms relied on the occurrence of at least one diagnosis. We used the Atlas tool to develop these algorithms, resulting in a total of ten distinct cohorts.
The cohorts were generated and evaluated across a range of data sources: including the claims-based databases Merative™ MarketScan® Commercial Claims and Encounters Database (CCAE), Optum Clinformatics® Data Mart (CDM) – Date of Death (Optum DOD), and the Merative™ MarketScan® Medicare Supplemental and Coordination of Benefits Database (Medicare). Additionally, general practitioner records from IQVIA® LPD in Australia and electronic health record (EHR) data from Optum® de-identified Electronic Health Record (Optum® EHR) and Premier Healthcare Database (Premier Premier) were utilized. ALL data sources contained measurement data from outpatient and/or inpatient encounters with at least partial coverage.

We assessed the agreement between the diagnosis-based and measurement-based phenotype algorithms by examining three key aspects: the proportion of patients identified solely by each approach, the proportion identified by both approaches, and the overlap of identified patients. The overlap represents the proportion of patients captured by both algorithms among those identified by the measurement-based algorithm. In addition, we conducted a comparison of covariate distributions among individuals who met each definition to evaluate the agreement in patient characteristics. We utilized the CohortDaignistic R package [3] to generate all the results.

Results

The detailed results are available at: https://data.ohdsi.org/Ohdsi2023AgreementMeasurementDiagnosis. Table 1 presents a heatmap depicting the overlap between diagnosis-based and measurement-based phenotype algorithms stratified by condition across each data source. Overall, a substantial heterogeneity in results was observed across data source and by condition. While CK and GFR measurements identified a relatively small number of patients of rhabdomyolysis and

end-stage renal disease respectively, a considerable number of patients was identified using measurement among the blood disorders, particularly thrombocytopenia. Of those, 3-75% were also identified by diagnosis.

Figure 1 illustrates the covariate distribution among patients diagnosed with thrombocytopenia, compared with those identified through measurement methods, on the index date in the IQVIA® LPD Australia dataset. A visible clustering of dots along the diagonal lines suggests a similarity in the clinical characteristics of patients between the two cohorts. Off-diagonal dots primarily represent measurements that were available in one cohort but not the other, which is a characteristic inherent to the design of the study. This pattern was consistent for all blood disorders across most data sources. However, for end-stage renal disease, the characteristics of patients identified by diagnosis was not comparable to those identified though measurement (figure 2).

Conclusion

Our findings suggest that, across the data sources examined, the utility of measurements in phenotyping varied by clinical condition and by data sources. Incorporating certain measurements alongside diagnosis could result in the identification of up to five times more patients for some conditions in some data sources. This substantial increase could significantly impact the sensitivity and specificity of the phenotype algorithm employed.

Additionally, our data revealed a comparable distribution of clinical characteristics among patients identified using either the measurement-based or diagnosis-only approaches among the blood disorder phenotypes. This similarity suggests both methods may be capturing the same clinical events, which suggests their combined use may increase sensitivity in patient identification in the examined data sources.

Our findings offer a framework for evaluating the utility of measurements for defining a phenotype within a given data source. This research can help investigators to more accurately and effectively phenotype patients, potentially leading to improve reliability of observational evidence. Empirically

References:

1. Khare Ritu, Kappelman Michael D, Samson Charles, et al. Development and evaluation of an EHR-based computable phenotype for identification of pediatric Crohn's disease patients in a National Pediatric Learning Health System. *Learning Health Systems*. 2020;4(4):e10243.
2. Gibson Teresa B, Nguyen Michael D, Burrell Timothy, et al. Electronic phenotyping of health outcomes of interest using a linked claims-electronic health record database: findings from a machine learning pilot project. *Journal of the American Medical Informatics Association*. 2021;28(7):1507-1517.
3. https://ohdsi.github.io/CohortDiagnostics/. Available at. Accessed.

Table and figures:

Table 1: The overlap between the diagnosis based and measurement-based phenotype algorithms by phenotype in each data source.

	Thrombocytopenia					Neutropenia				
Database Name	**D Only**	**M Only**	**Both**	**overlap**	**N**	**D Only**	**M Only**	**Both**	**overlap**	**N**
CCAE	88.4%	8.7%	2.9%	25.0%	1,209,577	86.3%	11.4%	2.2%	16.2%	1,226,238
IQVIA® Ambulatory EMR	14.4%	78.3%	7.3%	8.5%	1,075,889	50.5%	37.7%	11.8%	23.8%	573,655
IQVIA® LPD in Australia	17.7%	79.2%	3.1%	3.8%	5,244	100.0%	0	0	NA	1,805
Medicare	88.6%	6.3%	5.2%	45.2%	501,880	94.3%	3.7%	2.0%	34.5%	239,815
Optum's DOD	50.8%	29.5%	19.7%	40.0%	2,388,152	61.6%	27.5%	10.9%	28.4%	1,431,307
Optum® EHR	4.1%	72.6%	23.4%	24.4%	6,361,599	98.4%	0.8%	0.7%	46.7%	828,438
Premier	74.8%	17.8%	7.3%	29.1%	7,891,028	85.0%	10.8%	4.2%	28.0%	2,094,195
	Pancytopenia					**End-stage renal disease**				
	D Only	**M Only**	**Both**	**overlap**	**N**	**D Only**	**M Only**	**Both**	**overlap**	**N**
CCAE	97.7%	0.9%	1.5%	65.2%	234,617	100.0%	0.0%	<0.0%	NA	250,770
IQVIA® Ambulatory EMR	47.4%	31.1%	21.6%	50.0%	83,100	97.6%	0.9%	1.5%	62.5%	253,075
IQVIA® LPD in Australia	100.0%	0.0%	0.0%		84	39.6%	57.4%	3.0%	5.0%	197
Medicare	97.2%	0.7%	2.1%	75.0%	139,462	100.0%	0.0%	0	NA	223,704
Optum's DOD	85.4%	5.2%	9.4%	64.4%	406,264	99.0%	0.1%	0.9%	90.0%	527,668
Optum® EHR	24.5%	18.5%	57.0%	75.5%	433,971	100.0%	0.0%	0.0%	NA	341,753
Premier	88.4%	3.4%	8.2%	70.7%	1,527,731	100.0%	0.0%	0.0%	NA	2,700,277
	Rhabdomyolysis									
	D Only	**M Only**	**Both**	**overlap**	**N**					
CCAE	NA	NA	NA	NA	NA					
IQVIA® Ambulatory EMR	NA	NA	NA	NA	NA					
IQVIA® LPD in Australia	NA	NA	NA	NA	NA					
Medicare	NA	NA	NA	NA	NA					
Optum's DOD	NA	NA	NA	NA	NA					
Optum® EHR	99.5%	0.2%	0.3%	60.0%	161,254					
Premier	100.0%	0.0%	0.0%	NA	879,951					

D Only: Proportion of patients identified by diagnosis-based phenotype only among total patients identified by either approach
M Only: Proportion of patients identified by measurement-based phenotype only among total patients identified by either approach
Both: Proportion of patients identified by both approaches among total patients identified by either approach
Overlap: Proportion of patients identified in both approaches among those identified by the measurement-based algorithm

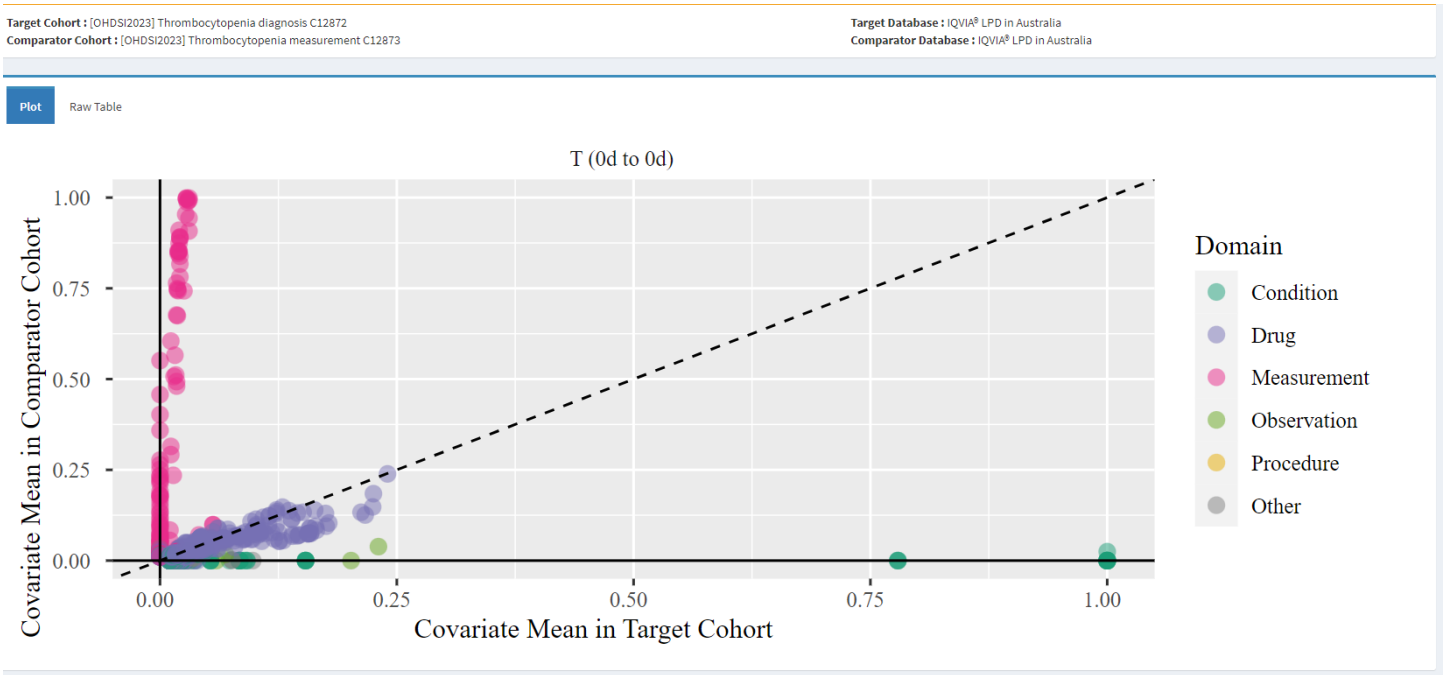

Figure 1: The covariate distribution among patients with thrombocytopenia identified through diagnosis compared to those identified through measurement on index date in IQVIA® LPD Australia

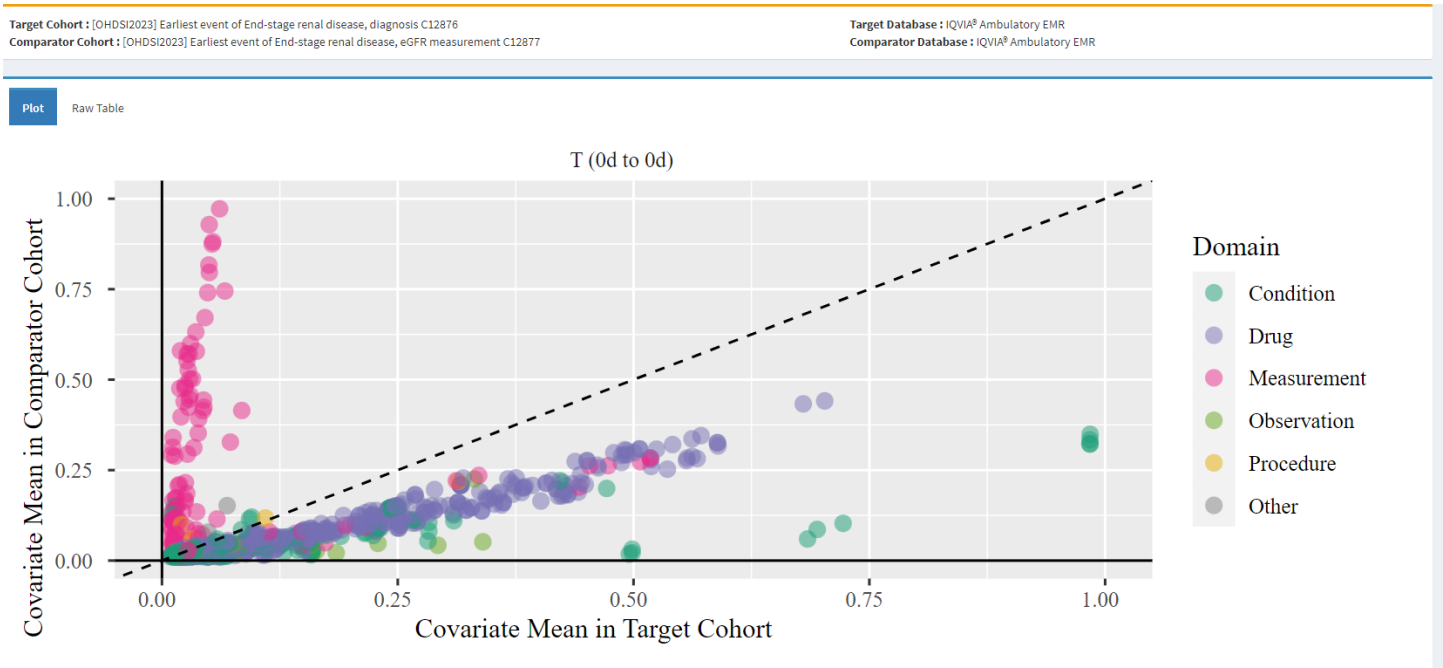

Figure 2 The covariate distribution among patients with End-stage renal disease identified through diagnosis compared to those identified through measurement on index date in IQVIA® Ambulatory EMR