

Argos Automatic Phenotype Generator for OMOP: Phenotypes for 1000+ Medical Conditions

Griffin M Weber, MD, PhD^{1,2}, Jeffrey Klann, PhD³, Shawn N. Murphy, MD, PhD^{3,4},
Aaron Abend, MBA⁵, Diane Keogh^{1,6}

¹Harvard Medical School, Boston, MA; ²Beth Israel Deaconess Medical Center, Boston, MA; ³Mass General Brigham, Boston, MA; ⁴University of Washington, Seattle, WA; ⁵Prognosis Data; ⁶i2b2 tranSMART Foundation

Background

A challenge with using electronic health record and medical claims data for research is that diagnosis codes often have low positive predictive value (PPV) for predicting the patient's true condition, or "phenotype". For example, about half of patients with an ICD-10 code for Type 2 Diabetes Mellitus (T2DM) do not actually have the disease. This might be because of data entry errors or because a patient was billed for a diagnostic test for the disease but does not necessarily have the condition.

This problem is addressed in OHDSI using phenotypes, which are standardized definitions for identifying patients with a medical condition. The "gold-standard" for creating phenotypes is an expert-derived consensus-building approach, such as through the eMERGE Network. However, this is a time consuming process, updating these phenotypes requires significant resources, and they are not easily adapted to the unique characteristics of an organization using them. As a result, the OHDSI community has been limited to only dozens of new phenotypes each year. Even if a phenotype exists for a study's target disease, the comorbidities, risk factors, and outcomes of a cohort, in most cases, still rely on the raw data.

Argos is a software platform that has recently been ported to OMOP that provides researchers with a simple user interface to explore clinical data and define cohorts. Argos includes algorithms that automatically construct weakly-supervised machine-learning based computed phenotypes. These are highly scalable, enabling an organization to generate more than 1,000 phenotypes using their OMOP data. These are probabilistic models, which are "tunable" to different levels of PPV and recall based on the use case. This Collaborator Showcase describes the Argos Automatic Phenotype Generator, its current adoption, and validation.

Methods

The computational pipeline in Argos combines two algorithms (**Figure 1**): (1) "KESER" uses co-occurrences of diagnoses, medications, and other concepts in the OMOP data to build embeddings, and then uses embedding regression to identify clinical features related to a medical condition [1]. For example, insulin and hypertension are associated with patients who have a diagnosis code for T2DM. (2) "KOMAP" then uses weakly-supervised machine learning to assign weights to these features to construct a phenotype model [2]. It incorporates the idea that patients who truly have a condition will have multiple occurrences of the diagnosis code, but normalizes this based on hospital utilization. A separate model is generated for each medical condition in the OMOP data, resulting in more than 1,000 phenotypes. Each phenotype model is run on every patient in the database to determine which phenotypes they have. The results are stored in the database and can be used through the Argos interface to define cohorts and explore their comorbidities.

The code is written in R and SQL and supports SQL Server and Oracle. It was originally developed for the i2b2 software platform, but can now run on OMOP through Argos. It is available as open-source code at <https://github.com/i2b2/i2b2-digitaltwin>.

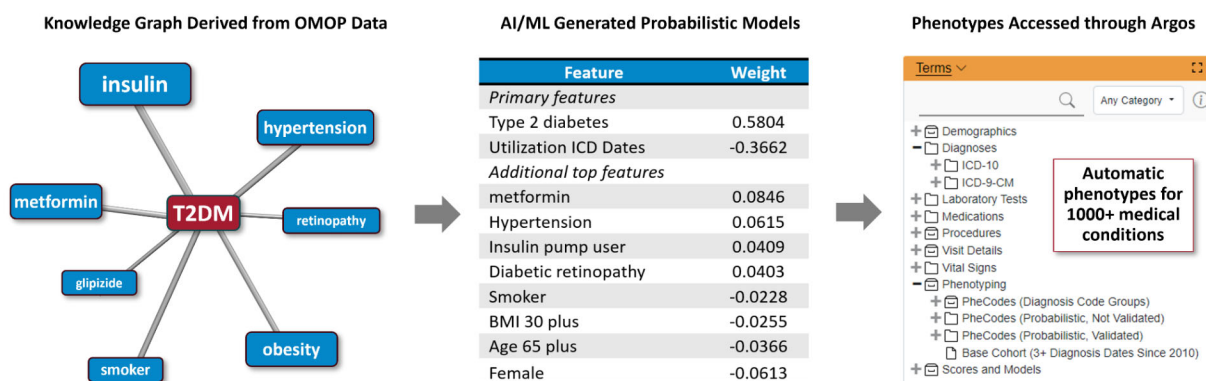


Figure 1. The Argos Automatic Phenotype Generator Pipeline. The KESER algorithm (left) creates a knowledge graph of clinical features related to medical conditions. KMOAP (middle) creates a machine learning model for each phenotype. The Argos UI (right) allows users to explore both the raw data and phenotypes of a cohort.

Results

At Beth Israel Deaconess Medical Center in Boston, Massachusetts, phenotypes were generated for 985 different medical conditions using about a week of processing time. At least one phenotype was assigned to 619,596 patients. A total of 6,952,005 patient-phenotype pairs were identified. A subset of 19 were validated using manual chart review from two clinical research nurses (**Figure 2**). On average, the default phenotype models had a PPV of 0.87 compared to a PPV of 0.57 when patients had a single ICD-10 diagnosis code for a condition and a PPV of 0.79 when patients had 2+ codes. The mean PPV of the phenotype models could be increased to 0.95, while keeping the recall above 0.5. Other institutions using this software include Mayo Clinic (generated phenotypes for 1,604 medical conditions and assigned these to 2,566,963 patients) and University of Florida (phenotypes for 1,186 conditions assigned to 1,082,522 patients).

Condition	1 ICD Code	1+ ICD Code	2+ ICD Codes	Computed Phenotype			Category	Inter-Rater Agreement
	PPV	PPV	PPV	Default PPV	Tuned PPV	Tuned Recall		
Cancer of prostate	1.00	0.93	0.94	0.94	0.93	1.00	1+ ICD Code	0.85
Breast cancer	0.10	0.75	0.97	0.96	0.90	0.90	2+ ICD Codes	0.83
Viral hepatitis	0.25	0.78	0.91	0.95	0.92	0.77	Computed Phenotype (high recall)	0.75
Major depressive disorder	0.82	0.80	0.88	0.83	0.92	0.69		
Chronic renal failure	0.22	0.48	0.75	1.00	0.93	0.68		
Type 2 diabetes	0.30	0.58	0.67	0.83	0.94	0.65		
Iron deficiency anemias	0.80	0.80	0.86	0.83	0.91	0.63		
Essential hypertension	0.33	0.73	0.83	1.00	0.90	0.62		
Gastroesophageal reflux disease	0.71	0.75	0.82	0.89	0.90	0.60		
Myocardial infarction	0.50	0.55	0.63	0.90	0.93	0.59		
Back pain	0.70	0.75	0.80	0.94	0.94	0.57		
Rheumatoid arthritis	0.67	0.75	0.85	0.85	1.00	0.53		
Type 1 diabetes	0.80	0.65	0.64	0.81	0.93	0.50	Computed Phenotype (low recall)	0.66
Asthma	0.73	0.78	0.83	1.00	0.92	0.35		
Sepsis	0.60	0.68	0.83	0.80	1.00	0.22		
Sprains and strains	0.50	0.68	0.79	0.75	1.00	0.22		
Acute upper respiratory infections	0.83	0.78	0.69	0.79	1.00	0.13		
Benign neoplasm of skin	0.38	0.43	0.50	0.50	1.00	0.12		
Skin cancer	0.50	0.70	0.79	0.86	1.00	0.07		
Mean:	0.57	0.70	0.79	0.87	0.95	0.52		

Figure 2. Manual chart review evaluation of 19 selected automatically generated phenotypes.

Conclusion

The Argos Automatic Phenotype Generator isn't perfect. Although it can reach high PPV for most phenotypes, it did not out-perform the raw diagnosis codes for a few conditions like prostate and breast cancer, and the recall was low in some others. The enormous benefit, though, is the scalability. This doesn't replace expert-driven phenotypes. However, it fills in the gap for all the medical conditions that do not yet have a "gold-standard" phenotype. It improves cohort identification for these conditions and provides more accurate comorbidity profiles than using the raw data. The automatic feature selection and model weights can also be used by experts as a starting point in building more robust models.

References

1. Hong C, Rush E, Liu M, et al. Clinical knowledge extraction via sparse embedding regression (KESER) with multi-center large scale electronic health record data. NPJ Digit Med. 2021 Oct 27;4(1):151. doi: 10.1038/s41746-021-00519-z. PMID: 34707226; PMCID: PMC8551205.
2. Liao KP, Sun J, Cai TA, et al. High-throughput multimodal automated phenotyping (MAP) with application to PheWAS. J Am Med Inform Assoc. 2019 Nov 1;26(11):1255-1262. doi: 10.1093/jamia/ocz066. PMID: 31613361; PMCID: PMC6798574.