Name:	Juan M. Banda, Alison Callahan, David Kale, Vladimir Polony, Nigam H. Shah.
Affiliation:	Stanford University, University of Southern California
Email:	jmbanda@stanford.edu
Presentation type (select one):	Collaborator Demonstration

Advanced Temporal Language Aided Search for the OHDSI community

Juan M. Banda, Ph.D.[1], Alison Callahan, Ph.D.[1], David Kale, M.S.[2] Vladimir Polony, M.S.[1] Nigam H. Shah, MBBS, Ph.D.[1]

[1]Stanford Center for Biomedical Informatics Research, Stanford, CA; [2] University of Southern California, Los Angeles, CA

Abstract

Observational research using electronic health records (EHR) typically starts with a phenotyping effort to define the cohort of patients to be studied. In the Observational Health Data Sciences and Informatics (OHDSI) community there are two complementary ways of defining phenotypes. The first is via implementing rule-based phenotype definitions from sources such as Phenotype Knowledge Base (PheKB), and the second is using Automated PHenotype Routine for Observational Definition, Identification, Training and Evaluation (APHRODITE), which produces a statistical phenotype model. In this demo we present a search engine, built on top of the Observational Medical Outcomes Partnership (OMOP) Common Data Model version 5, which allows rapid translation of consensus phenotype algorithms to define a phenotype in near real-time. We also present 29 definitions we have transcribed from PheKB definitions and are available via our search engine.

Introduction

Phenotyping algorithms using EHR data identify patients with a given condition and have led to new clinical discoveries[1,2]. Until recently, such algorithms mostly relied on rule-based definitions[3]. We have previously introduced APHRODITE to the OHDSI community, for creating statistical models for phenotypes[4]. In this demonstration we introduce a search engine based on Advanced Temporal Language Aided Search to supplement the OHDSI phenotyping efforts for rule-based definitions. This search engine consists of a patient centric data model, an in-memory database, and a structured language that supports temporal operations for near-real time phenotyping.

Advanced Temporal Language Aided Search language

The query language supports searching over patient features including gender, race and age, and over visit information including visit type, condition and procedure codes, medications prescribed, laboratory tests administered and their results. For example, to find females with a diagnosis of asthma, a query using the AND command finds patients that satisfy all criteria:

```
AND(GENDER="FEMALE", SNOMED=195967001)
```

Temporal operations in the query language allow queries for patients who had multiple events occur at the the same time, patients who had one event (diagnosis, procedure, medication or laboratory test) followed by another, patients with a history of a given event, and patients who never experienced an event. For example, to find patients who were prescribed morphine during an inpatient visit, a query using the INTERSECT command finds patients for whom the specified criteria occurred at the same time point:

```
INTERSECT(VISIT TYPE="inpatient visit", RX=7052)
```

To find patients who had a high HbA1C reading followed by a diagnosis of type 2 diabetes mellitus (T2DM), a query using the SEQUENCE command finds events of the specified type in a temporal sequence:

```
SEQUENCE(LAB("Hemoglobin A1c (Glycated)", "HIGH"), SNOMED=44054006*)
```

Temporal operations return the patients' time intervals for which the query condition was true. In the example query above, the asterisk indicates that the SEQUENCE command should return the time intervals for which the T2DM diagnosis code was present. It is also possible to search for events that occur within a given period of time of one another, such as the prescription of the antihypertensive drug atenolol within 30 days of a hypertension diagnosis:

```
SEQUENCE(SNOMED=59621000*, Rx=1202)+(-30 days, -1 day)
```

These types of temporal operations enable the execution of electronic phenotype definitions from sources such as PheKB as search engine queries, faster than would be possible using SQL.

Phenotype translation using Advanced Temporal Language Aided Search

In a single two-hour session, we used our search engine to implement several phenotype definitions from four standard sources: the Phenotype KnowledgeBase (PheKB)[5], the OMOP Health Outcomes of Interest (OMOP HOI) library[6], the Agency for Healthcare Research and Quality (AHRQ) Quality Indicator Resources[7], and the 2015 CMS Physician Quality Reporting System (CMS PQRS)[8]. We also implemented a subset of definitions used in recent published research[9]. In total, we implemented 29 phenotype definitions as queries in a span of roughly two hours, with an average of 1.93 definitions created per person in our team of 15 people. These included Type 2 Diabetes Mellitus and Atrial Fibrillation (from PheKB), Acute Myocardial Infarction and Acute Kidney Injury (OMOP HOI), Sepsis (AHRQ), Asthma (CMS PQRS), and Autism[8]. Table 1 summarizes the implemented definitions.

Of the 29 phenotypes implemented, 13 (45%) originated from PheKB, 8 (27.5%) from OMOP HOI, and 8 (27.5%) elsewhere. All but four definitions combined multiple data modalities (e.g., lab results, medication, procedures, text of clinical notes) with ICD-9 codes, and 14 definitions (48%) utilized multiple levels in the ICD9 hierarchy to determine sub-case types or to distinguish cases from controls. Thirteen definitions (45%) included explicit temporal criteria, such as searching for a sequence of elevated troponin levels followed by an ECG within 7 days to identify Acute Myocardial Infarction. Such criteria can be challenging and complex to express as SQL queries.

After initial implementation, we performed a careful manual review of the queries and confirmed that 22 (76%) definitions were fully consistent with the documented definitions and that the remaining seven required only minor revisions. This is noteworthy because most of our team had no experience with PheKB or OMOP HOI phenotype definitions prior to the working session.

Table 1. Summary of phenotypes implemented using Advanced Temporal Language Aided Search (ATLAS).

	PheKB	OMOP HOI	Other	Total
Total	*13*	*8*	*8*	*29*
Used lab results	3	2	0	*5*
Used medications	8	2	0	*10*
Used procedures	8	5	0	*13*
Used text mentions in clinical notes	8	2	3	*13*
Used explicit temporal criteria	5	5	3	*13*
Used multiple branches	11	3	0	*14*

Conclusion

Advanced Temporal Language Aided Search can allow rapid operationalization of expert created definitions, and enable instantaneous generation of labeled training data for phenotype models. The search engine, and the APHRODITE package, are designed to enable rapid electronic phenotyping for accelerating clinical research. We present 29 implemented expert-generated phenotype definitions and 3 statistical-model based phenotype definitions (APHRODITE models) to the community.

References

1. Ritchie MD, Verma SS, Hall MA, Goodloe RJ, Berg RL, Carrell DS, et al. Electronic medical records and genomics (eMERGE) network exploration in cataract: several new potential susceptibility loci. Mol Vis. 2014;20: 1281–1295.
2. Lin C, Karlson EW, Dligach D, Ramirez MP, Miller TA, Mo H et al. Automatic identification of methotrexate-induced liver toxicity in patients with rheumatoid arthritis from the electronic medical record. JAMIA. 2014.
3. Wright A, Pang J, Feblowitz JC, Maloney FL, Wilcox AR, Ramelson HZ, et al. A method and knowledge base for automated inference of patient problems from structured data in an electronic medical record. JAMIA. 2011;18(6):859–867.
4. Agarwal V, Podchiyska T, Banda, JM, Goel V, et al. Learning statistical models of phenotypes using noisy labeled training data. JAMIA 2016. DOI: http://dx.doi.org/10.1093/jamia/ocw028
5. Kirby JC, Speltz P, Rasmussen LV, Basford M, Gottesman O, Peissig PL, Pacheco JA, Tromp G, Pathak J, Carrell DS, Ellis SB, Lingren T, Thompson WK, Savova G, Haines J, Roden DM, Harris PA, Denny JC. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. J Am Med Inform Assoc. 2016 Mar 28; PMID: 27026615
6. Reich CG, Ryan PB, Schuemie MJ. Alternative outcome definitions and their effect on the performance of methods for observational outcome studies. Drug Saf. 2013. doi:10.1007/s40264-013-0111-1
7. AHRQ: http://www.qualityindicators.ahrq.gov/
8. CMS PQRS: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/
9. Lyalina S, Percha B, Lependu P, et al. Identifying phenotypic signatures of neuropsychiatric disorders from electronic medical records. J Am Med Inform Assoc 2013;20:e297–305.