

An Evaluation of the Impact of Vocabulary Evolution on Established Phenotypes

Frank DeFalco, Anthony Molinaro, Clair Blacketer

Background

Standard phenotypes are developed by defining cohort definitions that identify populations of interest for study. Cohort definitions rely on concept sets as part of inclusion and exclusion criteria definitions. Concept sets are lists of concepts in addition to logic settings for each concept. Logic settings provide a way to specify that descendants of a concept should be included, mapped concepts should be included, or rather that a concept and its descendant or mapped concepts should be excluded. Using these logic settings a concept set is resolved to a simple list of concepts when it is evaluated against a version of the Observational Medical Outcomes Partnership (OMOP) vocabulary. As the OMOP vocabulary is updated over time, the impact on existing concept sets and the cohort definitions that use them is not quantified.

Methods

An R script was developed to assess the changes introduced in cohort definitions and the resolution of their concept sets when changing between two different versions of the OMOP vocabulary. The script leverages the ROhdsiWebApi package and an installation of the WebAPI with two versions of vocabulary daemons configured. The script resolves all concept sets in a cohort definition using the two different versions of the OMOP vocabulary to produce two lists of concepts. The lists of concepts are then compared using set operations to check for additions and deletions of concepts between the two lists.

Results

Tabular results are generated to record the impact of the vocabulary update on the set of evaluated phenotypes. An overall summary table provides a listing of cohort definitions with a summary count of the number of concepts added or removed to the concept sets used in the definition as a result of the vocabulary change (Table 1).

Cohort Definition Id	Concepts Added	Concepts Removed
6700	16	0
6701	16	0
6705	179	1
6711	2	0
6712	10	0
6713	10	0
7832	6	0
7833	35	0

Table 1: Sample summary of overall phenotype library evaluation

A difference analysis report is generated providing listings of the modified concepts for each concept set in each cohort definition. The difference analysis report shows newly added concepts as a green highlighted row and concepts that were removed from the concept set resolution as a red highlighted row (Figure 1). The difference analysis report can assist in

interpreting the impact on a concept set and the cohort definition that utilizes it before requiring generation of all cohorts.

35804767	CAPIRI	Standard	14248	Regimen	HemOnc
35806400	Carboplatin and Etoposide (CE)	Standard	29282	Regimen	HemOnc
35806943	Carboplatin and Irinotecan	Standard	36182	Regimen	HemOnc
35803676	Cisplatin and Fluorouracil (CF)	Standard	3176	Regimen	HemOnc
35803585	Mitotane monotherapy	Standard	2619	Regimen	HemOnc
35804200	Neratinib and Paclitaxel	Standard	5782	Regimen	HemOnc
35804306	Neratinib monotherapy	Standard	9469	Regimen	HemOnc
2938218	{10 (Glucose 3560 MG / Potassium Chloride 300 MG / Sodium Chloride 450 MG / sodium citrate 530 MG Oral Powder) / 10 (loperamide hydrochloride 2 MG Oral Tablet) } Pack	Standard	OMOP5154890	Drug	RxNorm Extension
2934965	loperamide 2.15 MG	Standard	OMOP5151705	Drug	RxNorm Extension
2934964	loperamide 2.15 MG [Diarrhea Relief]	Standard	OMOP5151704	Drug	RxNorm Extension
2934963	loperamide 2.15 MG Oral Tablet	Standard	OMOP5151703	Drug	RxNorm Extension
2934961	loperamide 2.15 MG Oral Tablet Box of 16	Standard	OMOP5151701	Drug	RxNorm Extension

Figure 1: Screenshot of concept set change report

Conclusion

An evaluation of the impact of changes to the OMOP standard vocabulary allows researchers to anticipate and manage the potential impact to existing phenotypes. An entire library of phenotypes can be evaluated and prioritized for remediation if necessary to account for the changes introduced by changes in the vocabulary used to standardize an OMOP data source.

References

1. ROhdsiWebAPI [Internet]; Available from <https://github.com/OHDSI/ROhdsiWebApi>
2. Vocabulary-v5.0 [Internet]; Available from <https://github.com/OHDSI/Vocabulary-v5.0>
3. WebAPI [Internet]; Available from <https://github.com/OHDSI/WebApi>