HemOnc: A New Standard Vocabulary for Oncology Drug Regimen Representation in the OMOP Common Data Model

Dmitry Dymshyts[1], Christian G. Reich[2], Michael J. Gurley[3], Harry Hochheiser[4], Zachary H. Moldwin[5], Rimma Belenkaya[6], Andrew E. Williams[7], Peter C. Yang[8,9], Jeremy L. Warner[9,10*]

[1]Odysseus Data Services, Inc.
[2]IQVIA
[3]Northwestern University
[4]University of Pittsburgh
[5]University of Illinois at Chicago College of Pharmacy
[6]Memorial Sloan Kettering Cancer Center
[7]Tufts University
[8]Massachusetts General Hospital, Harvard Medical School
[9]HemOnc.org, LLC
[10]Vanderbilt University Medical Center

Is this the first time you have submitted your work to be displayed at any OHDSI Symposium?

Yes _____ No___X___

Background: Systematic application of observational data to the understanding of impacts of cancer treatments requires detailed information models to support meaningful comparisons between treatment regimens. Unfortunately, details of systemic therapies are scarce in registries and data warehouses, primarily due to the complex nature of the protocols and a lack of standardization.

Methods: Since 2011, HemOnc.org, LLC has been creating a curated and semi-structured website of oncology drug regimens, HemOnc.org. As of May 2019, there are a total of 904 content pages with >460,000 lines of content. In coordination with them, Observational Health Data Sciences and Informatics (OHDSI) Oncology Subgroup have transformed a substantial subset of this content into the OMOP Common Data Model, with bindings to multiple external vocabularies, e.g., RxNorm and RxNorm Extension (SNOMED and ATC links will be built soon).

Results: An OHDSI Oncology Episode extension was proposed to capture HemOnc oncology regimens as Episode events. Currently, there are >73,000 concepts and >177,000 relationships in the HemOnc source vocabulary; there are 24 classes and 36 relationship types.
In OMOP there were added 4,678 concepts of such classes: Component, Procedure, Regimen, Route, Regimen type, Component Class, Brand Name, Context; and 43,248 relationships.
Use cases include mapping of oncology drug regimens to a standardized vocabulary, mapping regimen acronyms and shorthand found in the natural language of clinical notes, and uncovering patterns of care at the practice level. Each Regimen has relationships to RxNorm or RxNorm Extension concepts, detailing the nature of each component's participation within a Regimen:

concept_class_id	relationship_name	count
Regimen	Has antineoplastic - RxNorm (HemOnc)	4021
Regimen	Has immunosuppressor - RxNorm (HemOnc)	154
Regimen	Has local therapy - RxNorm (HemOnc)	130
Regimen	Has supportive medication - RxNorm (HemOnc)	1001

Another important feature is that RxNorm antineoplastic concepts have new hierarchical classification branches, describing both clinical and chemical structure aspects, while existing ATC hierarchy lacks these details. This makes 3031 new CONCEPT_ANCESTOR entries with up to 7 levels of separation between RxNorm Ingredient and HemOnc classification concepts. In the future this hierarchy will be merged with ATC making one Standard hierarchy (Despite HemOnc hierarchy having a lot of novel concepts it has a duplicates with ATC as well).

Discussion: The new standard vocabulary for oncology drug regimens has been added to OHDSI Vocabulary, and as usual for OHDSI Vocabularies it has connection to the core terminologies (RxNorm and RxNorm Extension currently, SNOMED and ATC in the future). The HemOnc vocabulary represents the most extensive effort in the public domain to date intended to capture the structure of oncology drug regimens.

References

1. Warner JL, Dymshyts D, Reich CG, Gurley MJ, Hochheiser H, Moldwin ZH, Belenkaya R, Williams AE, Yang PC. A new standard vocabulary for chemotherapy regimen representation in the OMOP common data model. JBI-19-290R2. Approved for publication.
2. J.L. Warner, A.J. Cowan, A.C. Hall, P.C. Yang, HemOnc.org: A Collaborative Online Knowledge Platform for Oncology Professionals, J Oncol Pract. 11 (2015) e336-50. doi:10.1200/JOP.2014.001511.
3. G. Hripcsak, J.D. Duke, N.H. Shah, C.G. Reich, V. Huser, M.J. Schuemie, M.A. Suchard, R.W. Park, I.C.K. Wong, P.R. Rijnbeek, J. van der Lei, N. Pratt, G.N. Norén, Y.-C. Li, P.E. Stang, D. Madigan, P.B. Ryan, Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers, Stud Health Technol Inform. 216 (2015) 574–578.