

# Classification of RxNorm and RxNorm Extension Vaccine-related Terms in the Vaccine Ontology

Jie Zheng, PhD<sup>1†</sup>, Xingxian Li<sup>1†</sup>, Xumeng Zhang<sup>1</sup>, Warren Manuel<sup>2</sup>, Taiyu Lin<sup>3</sup>, Le Liu<sup>4</sup>, Rashmie Abeysinghe, PhD<sup>5</sup>, Joy Hu<sup>1</sup>, Yuping Zheng<sup>6</sup>, Katelyn Hur<sup>7</sup>, Anna He<sup>8</sup>, Yuanyi Pan, MD<sup>1</sup>, Yang Qi, PhD<sup>9</sup>, Alexander Davydov<sup>10</sup>, Anna Ostropolets, PhD<sup>10</sup>, Anna Maria Masci, PhD<sup>11</sup>, Junguk Hur, PhD<sup>12</sup>, Licong Cui, PhD<sup>2</sup>, Barry Smith, PhD<sup>13</sup>, Yongqun He, PhD<sup>1</sup>

<sup>1</sup>University of Michigan Medical School, Ann Arbor, MI, USA; <sup>2</sup>School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA; <sup>3</sup>SDU-ANU Joint Science College, Shandong University, Weihai, Shandong, China; <sup>4</sup>Marine College, Shandong University, Weihai, Shandong, China; <sup>5</sup>Department of Neurology, The University of Texas Health Science Center at Houston, TX, USA; <sup>6</sup>Chinese University of Hong Kong, Shenzhen, Guangdong, China; <sup>7</sup>Red River High School, Grand Forks, ND, USA; <sup>8</sup>Huron High School, Ann Arbor, MI, USA; <sup>9</sup>IQVIA, Inc., King of Prussia, PA, USA; <sup>10</sup>Odysseus Data Services, Inc., Cambridge, MA; <sup>11</sup>University of Texas MD Anderson Cancer Center, Houston, TX, USA; <sup>12</sup>University of North Dakota, Grand Forks, ND, USA; <sup>13</sup>University at Buffalo, Buffalo, NY, USA.

<sup>†</sup>Co-first authors

## Background

The OHDSI OMOP CDM<sup>1</sup> is a widely recognized open-science community data model, standardizing data from diverse clinical domains and sources to support robust and reliable analysis. To fully leverage the power of OMOP CDM, various terminologies and ontologies are utilized. As standard OHDSI vocabularies in the Drug domain, RxNorm<sup>2</sup> and RxNorm Extension<sup>3</sup> represent vaccines licensed in the United States and globally, respectively. The existing RxNorm/RxNorm Extension vaccine hierarchies mainly focus on ingredients and components.

As an Open Biomedical Ontologies (OBO) Foundry<sup>4</sup> library ontology, the Vaccine Ontology (VO)<sup>5</sup> represents licensed/authorized vaccines, vaccines used in clinical trials and research, vaccine components, and vaccine responses. In this study, we classified and integrated the RxNorm and RxNorm Extension vaccine-related terms in the VO with the goal of enhancing systematic classification and analysis of vaccine-related information for vaccine researchers.

## Methods

RxNorm and RxNorm Extension vaccine and vaccine ingredient terms were first extracted from the OHDSI vocabulary database. An automated approach<sup>6</sup> was adopted to identify identical terms by comparing the similarity of the extracted term names and VO term labels, which was followed by a manual review. When RxNorm and RxNorm Extension vaccine-related terms were not present in VO, they were added into VO according to a vaccine design pattern (ODP)<sup>7</sup>. This VO ODP was modified to include additional RxNorm vaccine attributes. The vaccines with their associated information were formatted according to the ODP into a ROBOT<sup>8</sup> template file (.csv) and then converted to OWL format using ROBOT's template function. The mappings between VO and RxNorm/RxNorm Extension terms are available at: [https://github.com/vaccineontology/VO/blob/master/src/templates/vo\\_RxNorm.csv](https://github.com/vaccineontology/VO/blob/master/src/templates/vo_RxNorm.csv).

## Results

We extracted a total of 7,440 vaccine and vaccine ingredient terms, including 2,051 terms from RxNorm (version 2023-07-03) and 5,389 terms from RxNorm Extension (version 2023-08-24). Our automated

approach identified 681 RxNorm and 2 RxNorm extension terms that exist in the VO (release 2024-01-03).

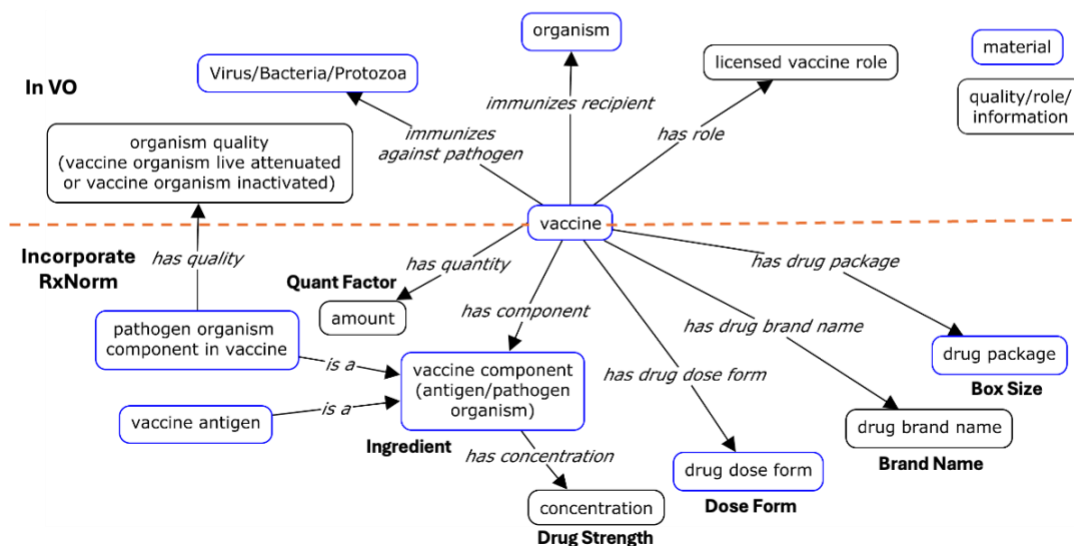
As illustrated in Figure 1, an updated VO ODP (Figure 1) was developed to semantically represent the RxNorm and RxNorm Extension vaccines together with their associated attributes and relations. Specifically, RxNorm includes six class types (i.e., clinical drug component, clinical drug, branded drug, quantified branded drug, and quantified branded box) and six attributes (i.e., ingredient, drug strength, dose form, brand name, quant factor, and box size). These are illustrated using the RxNorm term “1ML Hepatitis B surface antigen 0.01 MG/ML injection [H-B-VaxII] Box of 10” (Figure 1A).

In VO, the RxNorm ingredient is represented as a ‘vaccine component’ and other RxNorm class types are represented as vaccines. The existing VO ODP<sup>7</sup> focuses on vaccine attributes: 1) pathogen that a vaccine immunizes against; 2) a vaccine immunizes recipient; 3) vaccine component (e.g. ‘vaccine antigen’, ‘vaccine adjuvant’); 4) vaccine platform. We further extend the existing design pattern by representing specific RxNorm attributes (such as ‘dose form’ and ‘box size’) as ontology terms and linking the vaccine with these attribute-mapped ontology terms by specific relations (Figure 1B). The vaccine ingredients in RxNorm have two main types, antigens and pathogenic organisms, which could be mapped to more specific ‘vaccine component’ types, ‘vaccine antigen’ and ‘pathogen organism component in vaccine’ in VO. As an ingredient of the vaccine, pathogenic organisms are either inactivated or attenuated. VO contains organism quality, ‘vaccine organism live attenuated’ and ‘vaccine organism inactivated’, which was used to specify the status of a ‘pathogen organism component in vaccine’ (Figure 1B).

1ML Hepatitis B surface antigen 0.01 MG/ML injection [H-B-VaxII] Box of 10

Attributes	Quant Factor	Ingredient	Drug Strength	Dose Form	Brand Name	Box Size
<b>Classes</b>						
Ingredient		Hepatitis B Surface Antigen				
Clinical Drug Component		Hepatitis B Surface Antigen	0.01 MG/ML			
Clinical Drug		Hepatitis B Surface Antigen	0.01 MG/ML	Injection		
Branded Drug		Hepatitis B Surface Antigen	0.01 MG/ML	Injection	H-B-Vax II	
Quantified Branded Drug	1 ML	Hepatitis B Surface Antigen	0.01 MG/ML	Injection	H-B-Vax II	
Quantified Branded Box	1 ML	Hepatitis B Surface Antigen	0.01 MG/ML	Injection	H-B-Vax II	Box of 10

(A)



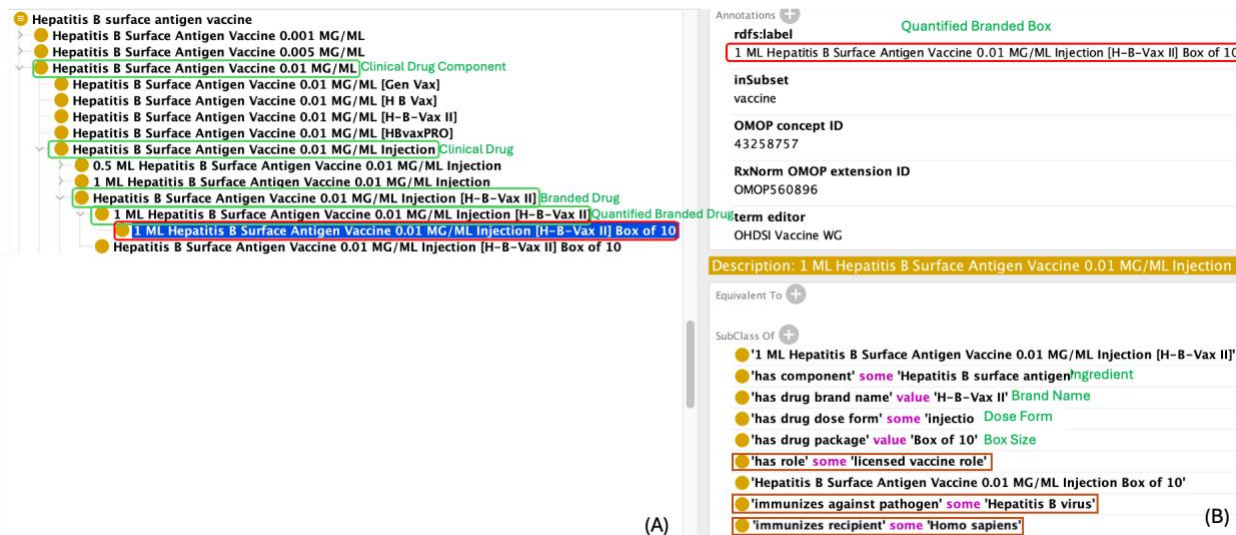
(B)

Figure 1. RxNorm-to-VO design illustration. (A) RxNorm vaccine representation design. (B) RxNorm-to-VO design pattern.

Figure 2 illustrates how the above VO design is used to represent the hierarchical structure of Hepatitis B surface antigen vaccines (Figure 2A) and one specific RxNorm vaccine in VO (Figure 2B). In addition to the RxNorm axioms, VO includes other logical axioms:

- 'vaccine immunizes against pathogen' some 'Hepatitis B virus'
- 'vaccine immunizes recipient' some 'Homo sapiens'
- 'has role' some 'licensed vaccine role'.

These axioms represent additional vaccine attributes that support computer-assisted vaccine classification.



**Figure 2. VO representation of a vaccine (highlighted in red box) in a hierarchical structure (A) and ontology representation (B).** Green boxes highlight the concepts in RxNorm or RxNorm Extension and the text in green indicates the mapped RxNorm concept class types or attributes. Orange boxes are new axioms added in VO based on the VO design pattern.

Our work resulted in the addition of 1,370 RxNorm and 4,089 RxNorm Extension vaccine-related terms to VO (release 2024-06-26). Seven pairs of RxNorm Extension vaccine terms were found having the same names and considered as the same vaccines. The terms were merged in VO by assigning 7 unique VO IDs. A total of 1,298 RxNorm Extension terms have incomplete names (containing '...' in long-length names); as a result, these terms have not yet been added to VO, and we are working to solve this issue.

Ontological representation of vaccines using logical axioms in VO enables us to construct the RxNorm and RxNorm Extension vaccines hierarchy based on vaccine attributes in addition to asserted hierarchy structures. The intermediate terms were added to group RxNorm and RxNorm Extension vaccine-related terms based on specific vaccine properties. With the addition of vaccine-against-pathogen information, we identified that RxNorm vaccines targeted 24 different pathogens and organized them according to the targeted pathogens in VO (Figure 3A). Figure 3B showed the hierarchy of influenza vaccines based on vaccine targeted flu season.

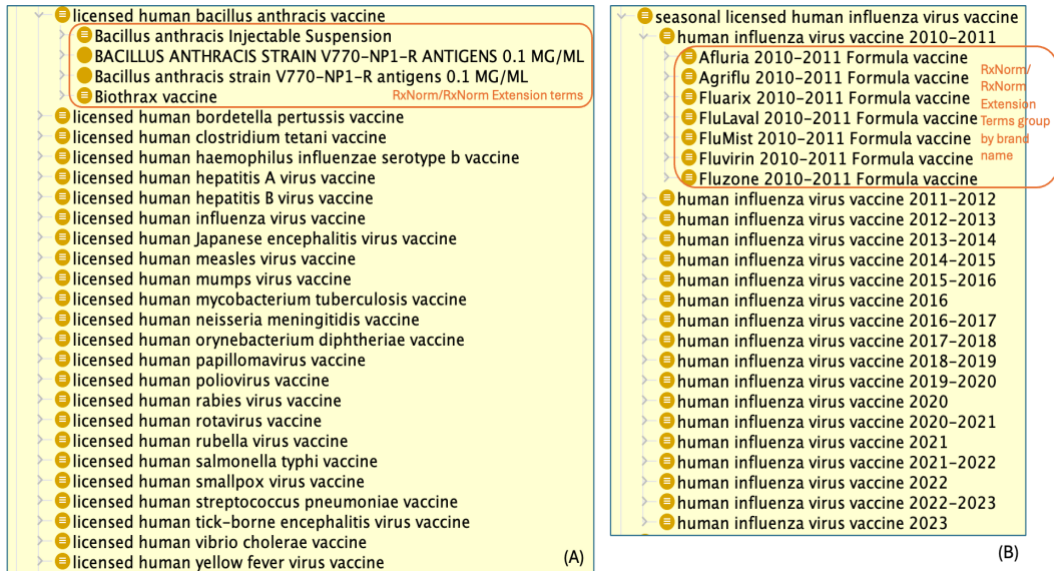


Figure 3. Hierarchy of RxNorm/RxNorm Extension vaccines based (A) vaccine against pathogen and (B) flu season.

Using a Description Logic (DL) query, we could retrieve vaccines of interest according to logical axioms. Figure 4 demonstrates a Description Logic (DL) query was developed to identify 315 vaccines against pathogen 'Hepatitis A' that have the dose form of 'injectable product'.

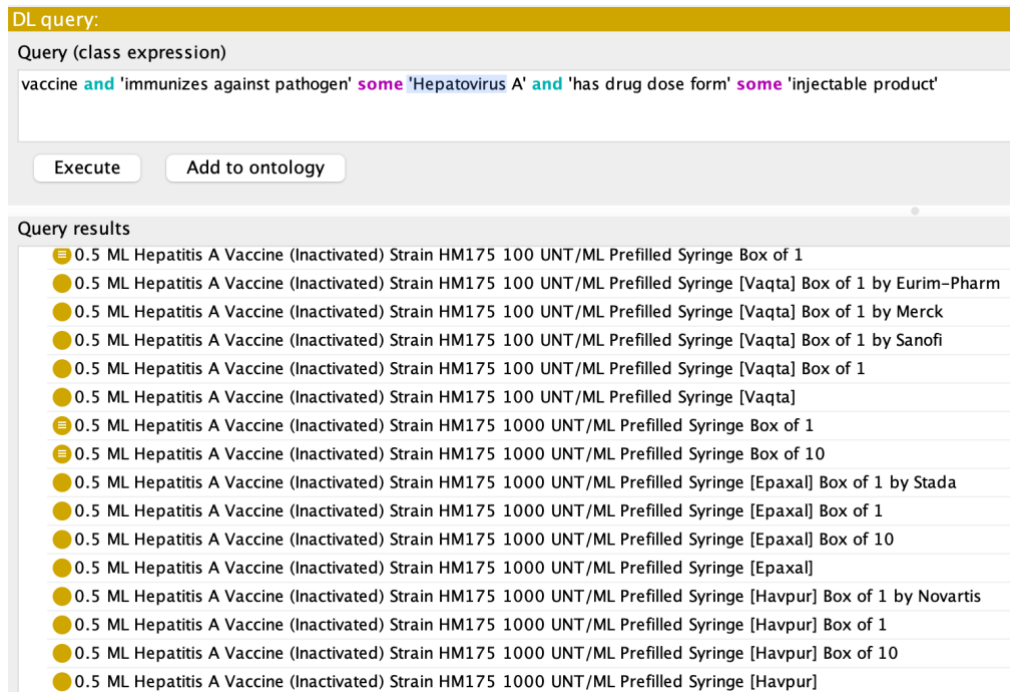


Figure 4. Demonstration of a DL-query based on the VO representation.

## Conclusion

We integrated 2,051 RxNorm terms (including vaccines and vaccine ingredients) and 4,091 RxNorm

Extension terms into VO, incorporating additional semantic relations. RxNorm attributes such as ingredient, dose form, brand name, and box size have been represented using logical axioms in VO. The newly added hierarchy aligns with the existing VO structure, and intermediate VO terms were added to provide new hierarchies of RxNorm and RxNorm Extension vaccine-related terms based on specific vaccine types such as targeted pathogen and disease types and flu seasons. This integration greatly enhances the representation, querying, and analysis of various vaccine information within a single ontological knowledge resource. A future task will be to identify the needs of OMOP and better support vaccine data analysis within the OMOP framework.

### References

1. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for observational researchers. *Stud Health Technol Inform.* 2015;216:574–8.
2. RxNorm. 2004; Available from: <https://www.nlm.nih.gov/research/umls/rxnorm/index.html>
3. Observational health data sciences and informatics. Ohdsi.org. Available from: [https://www.ohdsi.org/web/wiki/doku.php?id=documentation:international\\_drugs](https://www.ohdsi.org/web/wiki/doku.php?id=documentation:international_drugs)
4. Smith B, The OBI Consortium, Ashburner M, Rosse C, Bard J, Bug W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol.* 2007;25(11):1251–5. Available from: <http://dx.doi.org/10.1038/nbt1346>
5. He Y, Chen F, Scheuermann R, Todd T, Brinkman R, Colby L, et al. VO: Vaccine Ontology. *Derm Helv.* 2009;1–1. Available from: <https://www.nature.com/articles/npre.2009.3553.1>
6. Pan Y, Manuel W, Abeyasinghe R, Hao X, Davydov A, Yang Q, Lin AY, Cui L, He Y. Harmonization of OMOP vaccine-related vocabularies through the Vaccine Ontology. The 2023 OHDSI Global Symposium, Oct. 20-22, 2023, East Brunswick, New Jersey, USA.
7. Zheng J, Li X, Masci AM. Empowering standardization of cancer vaccines through ontology: enhanced modeling and data analysis. *J Biomed Semant.* 2024;15.
8. Jackson RC, Balhoff JP, Douglass E, Harris NL, Mungall CJ, Overton JA. ROBOT: A tool for automating ontology workflows. *BMC Bioinformatics.* 2019;20(1):407. Available from: <http://dx.doi.org/10.1186/s12859-019-3002-3>