

Vasculitis without phlebitis phenotype development using real-world data: development and evaluation study

Jill Hardin^{1,2}, Amir Sarayani³, Dina Gifkins¹, Tara Beaulieu¹, James Gilbert^{1,2}, Joel Swerdel^{1,2}

¹Janssen Research and Development, LLC, Titusville, NJ, USA

²Observational Health Data Sciences and Informatics (OHDSI), New York, NY

³Janssen Research and Development, LLC, Brisbane, CA, USA

Background

Vasculitis is a potentially debilitating, chronic, recurring inflammatory disease of the vascular system. Observational databases provide opportunities to study the epidemiology of vasculitis without phlebitis.

Objective

This study's objective was to develop a phenotype algorithm for vasculitis without phlebitis suitable for epidemiological studies based on a network of observational databases.

Methods

A data-driven approach was used to develop and evaluate the performance metrics of a vasculitis without phlebitis phenotype algorithm. A literature search identified prior vasculitis algorithms. Standardized databases from the Observational Medical Outcomes Partnership (1) version 5.3.1 (n=5) were used to develop the vasculitis without phlebitis phenotype algorithm excluding skin infections (SNOMED concept 4029043 and all descendants) in the 90 days prior to index. The PHOEBE (2), a tool in the Observational Health Data Sciences and Informatics (OHDSI) tool stack, was used to construct an initial concept set and it was refined after review by clinical and epidemiology experts. A phenotype cohort was defined in five nationwide United States (US) health insurance claims databases (see Table 1 for database descriptions) by requiring at least one diagnosis code of vasculitis, without phlebitis. Three open-source OHDSI diagnostic tools, ATLAS (3), CohortDiagnostics (4), and PheValuator (5), were used to develop the phenotype and evaluate performance metric estimates, including sensitivity, specificity, positive predictive value (PPV), and negative predictive value using a probabilistic approach.

Results

The final concept set expression had 178 concepts for vasculitis without phlebitis from the condition (n=173) and observation (n=5) domains, respectively. Identification of vasculitis without phlebitis ranged from 101,000 cases in Merative Medicare to 361,000 in Optum Clinformatics® DOD. Between 22% (IQVIA Pharmetrics) to 35% (Merative MarketScan® CCAE) of subjects had the code for "arteritis" (standard concept code: 314659), and between 11% (Merative CCAE) to 39% (Merative MarketScan® Medicare) had the code "temporal arteritis" (standard concept code: 4290976). The annual incidence rate estimates among subjects were stable across the databases (Figure 1). Rates were ~ 3 times higher in ages > 50 years compared to ages < 50 years. Rates were ~30% higher in females compared to males. The PheValuator tool estimated the mean sensitivity of the phenotype to be 77% and the mean positive predictive value to be 77% across the databases.

Conclusion

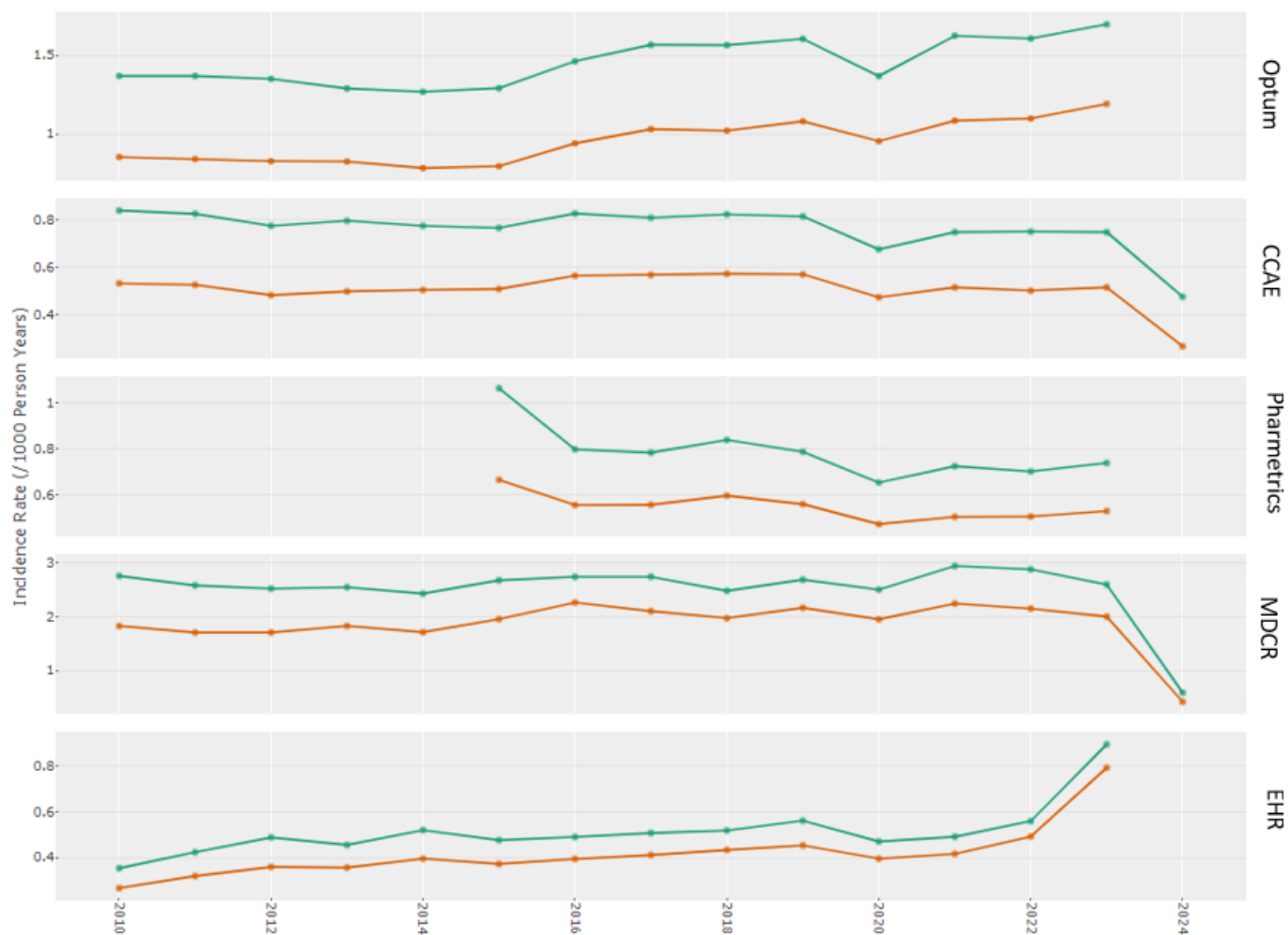
We developed a computable phenotype to measure vasculitis without phlebitis and provided performance metrics. Future studies can explore probabilistic algorithms rather than rule-based

approaches to improve the performance metrics for this phenotype.

Table 1: Description of Databases used in the study

Name (Abbreviation)	Years	Country	Data Type	Clinical Visits included	Number of Persons (millions)	Average Age at First Observation	Percent Female	Median Length of Follow-up (years)
Merative® MarketScan Commercial Claims and Encounters (CCAЕ)	2000-2024	US	Insurance Claims	Inpatient/outpatient	172	31	51	2.84
Merative® MarketScan Medicare Supplemental (MDCR)	2000-2024	US	Insurance Claims	Inpatient/outpatient	11	71	56	3.98
Optum’s Clinformatics® Data Mart - Date of Death (Optum)	2000-2023	US	Insurance Claims	Inpatient/outpatient	99	36	51	3.21
Optum® Electronic Health Record dataset (Optum EHR)	2007-2024	US	Electronic health records	Inpatient/outpatient	114	37	53	4.91
IQVIA® Pharmedics	2015-2023	US	Insurance Claims	Inpatient/outpatient	163	34	50	2.62

Figure 1. Incidence rate pattern over time generated by Cohort Diagnostics



(Optum: Optum Clinformatics® Data Mart - Date of Death; CCAE: Merative Commercial Claims; Pharmetrics: MDCR - Merative Medicare Supplemental; EHR – Optum Clinformatics® de-identified Electronic Health Record databases; data year was complete through 2023 for all databases)

Table 2- Summary of phenotype performance metrics estimated via PheValuator tool

Database Name	PheValuator Sensitivity	PheValuator PPV	PheValuator specificity	PheValuator NPV
Merative CCAE	0.804 (0.791 - 0.817)	0.775 (0.761 - 0.788)	1.000 (1.000 - 1.000)	1.000 (1.000 - 1.000)
Merative MDCR	0.535 (0.528 - 0.542)	0.858 (0.851 - 0.864)	0.999 (0.999 - 0.999)	0.996 (0.995 - 0.996)
Optum EHR	0.839 (0.827 - 0.850)	0.704 (0.691 - 0.717)	0.999 (0.999 - 0.999)	1.000 (1.000 - 1.000)
Pharmetrics	0.898 (0.888 - 0.908)	0.741 (0.728 - 0.754)	0.999 (0.999 - 0.999)	1.000 (1.000 - 1.000)
Optum DOD	0.781 (0.772 - 0.790)	0.809 (0.800 - 0.818)	0.999 (0.999 - 0.999)	0.999 (0.999 - 0.999)

References

1. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc.* 2015;22(3):553-64.
2. Ostropolets A, Ryan P, Hripcsak G. Phenotyping in distributed data networks: selecting the right codes for the right patients. *AMIA Annu Symp Proc.* 2023 Apr 29;2022:826-835.
3. <https://github.com/OHDSI/Atlas>
4. <https://github.com/OHDSI/CohortDiagnostics>
5. Swerdel JN, Schuemie M, Murray G, Ryan PB. PheValuator 2.0: Methodological improvements for the PheValuator approach to semi-automated phenotype algorithm evaluation. *J Biomed Inform.* 2022 Nov;135:104177.