Title: Development of Phenotype Algorithms and Characterizations of Primary Open-Angle Glaucoma Using Real-World Data

Authors: Nathan Hall^{1,2}, Rupa Makadia^{1,2}

¹Janssen Research and Development, Titusville, NJ; ²Observational Health Data Sciences and Informatics (OHDSI), New York, NY

Background: Glaucoma is the second-leading cause of blindness in the world and presents a serious public health problem. There are several subtypes of glaucoma, but primary open-angle glaucoma (POAG) is the most common, comprising approximately 80% of all glaucoma cases and affecting upwards of 2.2% of the world's population.^{1,2} Additionally, disease burden and risk of POAG tends to vary greatly across different demographic, geographical, and clinical factors.³ Given the potential for blindness to develop in POAG patients, varying levels of risk across different patient factors, and since wide-scale phenotypes of this disease do not exist yet, phenotyping and characterizing the presentation and risk factors for POAG is critical.

Methods: Six different phenotype algorithms were developed to identify six cohorts (C1-C6) of POAG patients after an extensive literature review of the condition. Data from five standardized databases (DB) within the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) were utilized.⁴ The DBs used in this study include the Optum© De-Identified Clinformatics® Data Mart Database - Date of Death-(Optum DoD), the IBM MarketScan® Medicare Database (IBM MDCR), the IBM MarketScan Commercial Claims and Encounters (IBM CCAE), the Optum® de-identified Electronic Health Record dataset (Optum EHR), and the Japan Medical Data Center (JMDC) database. Standardized difference plots were used to compare differences in prevalence of covariates across cohorts. Two open-source tools from the Observational Health Data Sciences and Informatics (OHDSI) program were used to create the phenotype algorithms and to run diagnostics on the resulting cohorts: ATLAS and CohortDiagnostics, respectively.^{5,6}

Results: The algorithmic definitions and cohort counts of unique POAG patient phenotypes for each of the six cohorts within the five DBs are outlined in **Table 1**.

Table 1. Unique primary open-angle glaucoma (POAG) patients by cohort specification and database.

Cohort Specifications		Database & POAG Patient Counts				
Cohort Name	Phenotype Algorithm	Optma DoD (v2012)	IBM CCAE (v2008)	IBM MDCR (72909)	Optum EHR (v1990)	JMDC (v2019
Colori 1 (C1)	Barlest recorded condition occurrence of POAG.	1,213,558	705,245	724,077	215,301	126,929
Celear 2 (C2)	Endast recorded condition accurrence of PDAG for patients with at lowe 1 years of prior continuous pharework and 2 PDAG diagnoses (including indice).	540,828	210,597	297,072	120,694	75,563
Colort 7 (C3)	Furthers excerted condition accurrence of POAG- recluding patients with a condition eccurrence of eccendary glassense my line prior to italics.	1,205,128	0672891	769,577	201,192	125,558
Colort 4 (C1)	Forkers excerted condition accumutes of glassense (any type)	3,608,543	2,085,526	1,578,378	1,055,901	738,758
Cohort 5 (C5)	Barbas recential condition securement of glucoma (any topy), real-biling satisms with (1) a condition occurrence of secondary glucoma, and (2) systemic steroid use, any time on or prior to index.	3,431,029	3.621.563	1,473,541	1462,773	727.731
Colast 6 (C.9)	Darkest recorded condition accurrence of open angle gimmonia (any type)	1,368,553	121,912	813,547	254,673	155,670

Across all DBs, the definition for C4 yielded the most patients, while the definition for C2 yielded the least. Within Optum DoD in for C3, there were 1,205,128 unique patients, with a

mean age of 70.4 and 3.4 years of median observation time in cohort. Female patients made up 56.2% of the cohort, and 5.9% of patients had a myopia diagnosis prior to index.

Comparisons of cohort characterizations using standardized differences between covariates revealed that earliest-recorded POAG (C1), earliest-recorded POAG excluding those with a secondary glaucoma diagnosis prior to index (C3), and earliest-recorded open-angle glaucoma (C6) patients exhibited little differences between each other, whereas all other comparisons highlighted stark differences in recurrent covariates, such as borderline open angle glaucoma diagnoses and antiglaucoma medication usage. The comparison between the most plausible cohort candidates (C3 and C5) for the strongest phenotype for POAG was profound (**Figure 1**).

Figure 1. Comparison between proportion of subjects in the C3 cohort description compared to the C5 cohort description for the Optum DoD dataset across different covariates. Points closest to the diagonal indicate similar proportions between the comparators.



Conclusions: This study uses real-world data and OHDSI tools to characterize the presentation of primary open-angle glaucoma across several databases using a range of phenotyping algorithms. Given the subtle and gradual presentation of this disease it can be difficult to develop an accurate phenotype, but the diagnostic results from comparing our cohort characterizations provide insights into differentiating factors between primary open-angle glaucoma and other, less common types of glaucoma. Generalized cohort definitions with less stringent algorithms result in larger sample sizes and capture more general glaucoma patients than stricter definitions. However, to develop a phenotype for primary open-angle glaucoma that is both accurate and robust, careful cohort selection based on specified glaucoma type and excluding potential underlying causes to glaucoma such as steroid usage is also critical. Our findings demonstrate that an ideal balance of specificity and sensitivity may be achieved when including pieces from both a general glaucoma and a primary open-angle glaucoma cohort definition together, to encapsulate true primary open-angle patients via glaucoma medication usage and those that have borderline and/or suspect glaucoma diagnoses. Further analysis using these tools, in addition to conducting an in-depth analysis using PheValuator⁸ to determine sensitivity, specificity, positive predictive value and negative predictive value is warranted for validation purposes. These additional analyses, in combination with considering more patient-level factors such as comorbidities, race, and other known risk factors for primary open-angle glaucoma will serve to hone the phenotype algorithm and ultimately provide hope to the millions of patients suffering from this debilitating disease.

References:

- 1. Allison, Karen et al. "Epidemiology of Glaucoma: The Past, Present, and Predictions for the Future." *Cureus* vol. 12,11 e11686. 24 Nov. 2020, doi:10.7759/cureus.11686.
- 2. Sugiyama K. (2012). "A challenge to primary open-angle glaucoma including normalpressure. Clinical problems and their scientific solution". *Nippon Ganka Gakkai zasshi*, 116(3), 233–268.
- 3. Zhang, N., Wang, J., Li, Y. et al. Prevalence of primary open angle glaucoma in the last 20 years: a meta-analysis and systematic review. *Sci Rep* 11, 13762 (2021). https://doi.org/10.1038/s41598-021-92971-w.
- 4. Christian Reich, et al. OHDSI/CommonDataModel. 2018 [cited 2022 06/13/2022]; Available from: https://ohdsi.github.io/CommonDataModel/.
- 5. GitHub Repository; Available from: https://github.com/OHDSI/Atlas.
- 6. GitHub Repository; Available from: <u>https://github.com/OHDSI/CohortDiagnostics</u>.
- 7. Swerdel JN, Hripcsak G, Ryan PB. PheValuator: Development and evaluation of a phenotype algorithm evaluator. Journal of biomedical informatics. 2019;97:103258.