

# A Natural Language Processing Approach to Identify Patients with Uveitic Macular Edema in the IRIS® Registry

Peng Jin, PhD<sup>1</sup>, Marie Humbert-Droz, PhD<sup>1</sup>, Kristian Garcia, MPH<sup>1</sup>, Helene Fevrier, MPH<sup>1</sup>, Durga Borkar, MD, MMCI<sup>1</sup>, Abhishek A. Nair, PhD, MS<sup>2</sup>, David Harrison, PhD, MA<sup>2</sup>, Zhongdi Chu, PhD, MSc<sup>1</sup>

1. Verana Health, CA, United States; 2. Bausch & Lomb Americas Inc., NJ, United States

## PURPOSE

- Uveitic Macular Edema (UME) patients are difficult to identify in real-world clinical research settings using structured data.
- To develop a natural language processing (NLP) algorithm to identify patients with active uveitic macular edema (UME) from electronic health records (EHR) data in the American Academy of Ophthalmology IRIS® Registry (Intelligent Research

## METHODS

- Fellowship-trained retina specialists defined a combination of ICD-10 codes for macular edema (ME) and non-infectious uveitis in structured data, and a list of UME keywords for non-historical ME in association with non-infectious uveitis in unstructured data to identify patients with UME.
- A heuristic NLP algorithm was developed and validated to identify patients with an active UME diagnosis at a given encounter based on the unstructured data definition using a SpaCy PhraseMatcher.
- Finally, the proposed NLP algorithm was used to identify patients with active UME in the IRIS Registry. The number of UME patients identified by the proposed NLP algorithm was compared to the number identified based on the ICD-10 codes alone.

## METHODS (cont.)

### Algorithm Development

- Notes from 500 randomly selected patients with UME keywords in their clinical records were labeled by clinical experts to determine their UME status: active UME or no/unknown active UME.
  - Patients were randomly sampled with stratification on EHR systems.
  - Surfaced notes were truncated and concatenated at patient-encounter level for labeling purpose.
  - At least two trained clinical experts participated in the labeling process to achieve inter-rater reliability.
- This labeled dataset was split 7:3 for algorithm development (n = 350) and validation (n = 150), with stratification on EHR systems and labels.
- The NLP algorithm was built using SpaCy PhraseMatcher module and business rules based on the development set.

### Algorithm Validation

- The proposed NLP algorithm was evaluated on the validation set using accuracy, sensitivity and specificity as the performance metrics.
- To validate data quality and ensure the outputs met clinical expectations, the final algorithm was applied to all patients with UME keywords in their clinical notes from the IRIS® Registry to identify their UME status at encounter level.

## RESULTS

### Algorithm Performance

- There were 231,543 patients with UME keywords in their clinical records using IRIS Registry data from January 1st, 2016 to August 16th, 2023.
- The algorithm achieved an accuracy, sensitivity and specificity of 0.83, 0.95 and 0.73, respectively, using the validation set (n = 150).
- The F1-scores for "Confirmed active UME" and "No/unknown active UME" were 0.81 and 0.83, respectively.

## RESULTS (cont.)

**Table 1.** Performance Metrics on the Validation Set (n = 150)

	N	Precision	Recall	F1-score
<b>Confirmed active UME</b>	60	0.70	0.95	0.81
<b>No/unknown active UME</b>	90	0.96	0.73	0.83

### Active UME Patients Identified by the NLP Algorithm

- The NLP algorithm was able to determine patients with "active UME" or "no/unknown active UME" at any give encounter.
- Out of 231,543 patients with UME keywords in their clinical records, 129,316 patients were confirmed with active UME at the encounter level by the proposed NLP algorithm.

**Table 2.** Number of Patients with Active or No/Unknown UME Status Identified by the NLP Algorithm

	Number of Patients
<b>Patients with UME keywords in clinical notes*</b>	231,543
Patients with "active UME" at any give encounter	129,316
Patients with "no/unknown active UME" at any give encounter	142,221

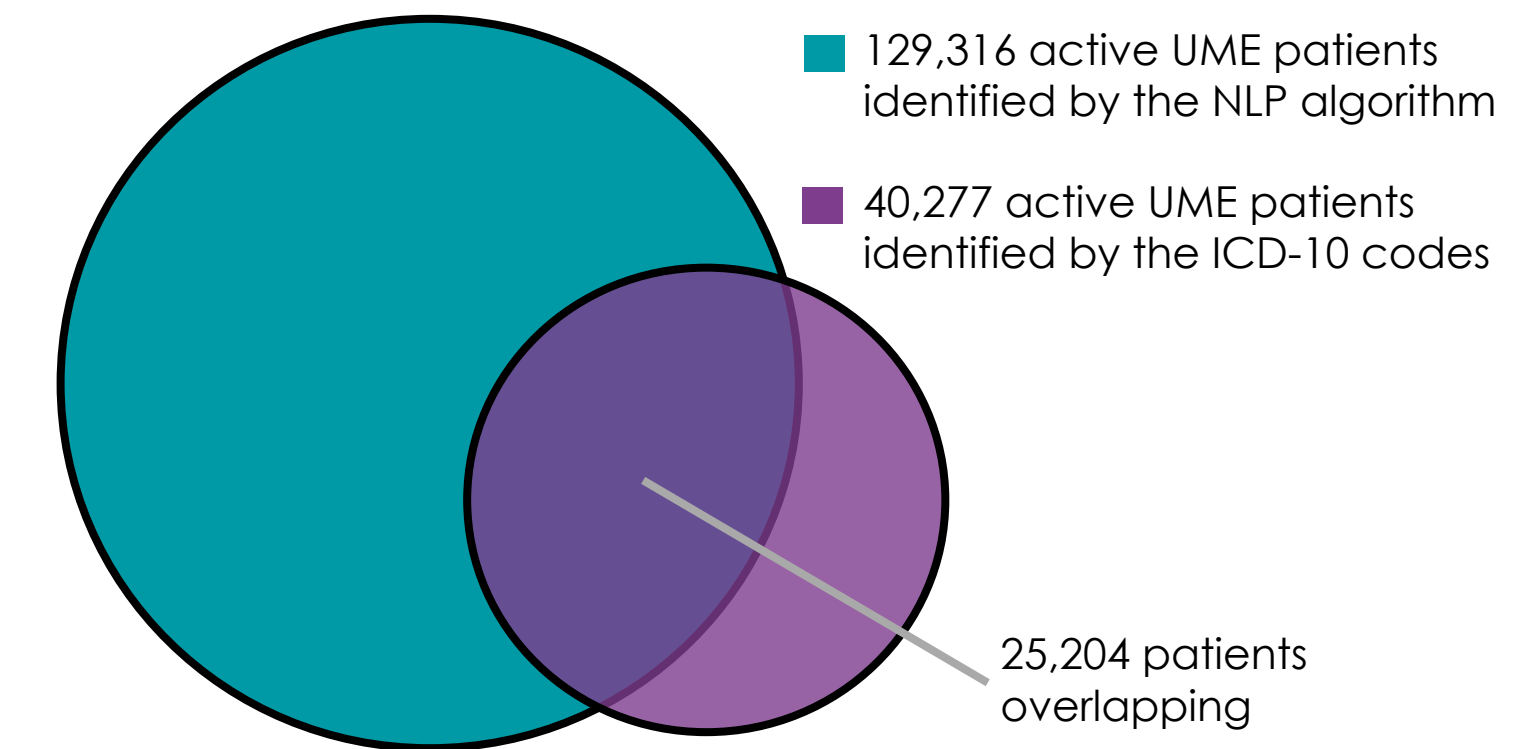
\*IRIS Registry data from January 1st, 2016 to August 16th, 2023.

### Comparison against Structured Data

- Patients with active UME that were identified by the NLP algorithm were compared to those identified by ICD-10 codes only.
- Out of 40,277 patients that were identified as having active UME diagnosis using ICD-10 codes, 25,204 patients were also confirmed with active UME at least one encounter by the proposed NLP algorithm.

## RESULTS (cont.)

**Figure 2.** Number of Patients with Active UME by approach



## CONCLUSIONS

**3X** NLP algorithm identified **three times** more patients with **active UME** compared to only using ICD-10 codes

**Target** NLP algorithm achieved a **high accuracy, and sensitivity and moderate specificity** in identifying active UME patients

Heuristic NLP approach is an enhanced solution to conduct real-world evidence studies in the UME patient population

## REFERENCES

McKay, K. Matthew, et al. "Assessing the uniformity of uveitis clinical concepts and associated ICD-10 codes across health care systems sharing the same electronic health records system." JAMA ophthalmology 139.8 (2021): 887-894.

Lee, Cecilia S., et al. "American Academy of Ophthalmology Intelligent Research in Sight (IRIS®) registry and the IRIS registry analytic center consortium." Ophthalmology Science 2.1 (2022).

## DISCLOSURE

This study was funded by Bausch & Lomb Americas Inc. Abhishek Nair and David Harrison are employees of Bausch & Lomb Americas Inc. Peng Jin, Marie Humbert-Droz, Helene Fevrier and Zhongdi Chu are employees of Verana Health. Durga Borkar is a consultant of Verana Health, AbbVie, Apellis, Glaukos, Genentech, and Iveric Bio.



**Figure 1.** Workflow of the Proposed NLP Algorithm

