Identifying Risk Profiles for Early Treatment Discontinuation in Geographic Atrophy Using Machine Learning and SHAP Clustering

MSR126

Ashwin Rai,[1] Ariel Berger,[1] Andre Ng,[2] Victoria Ikoro,[3] Devika Bhandary[2]

[1]Thermo Fisher Scientific, Waltham, MA, USA; [2]Thermo Fisher Scientific, London, UK; [3]Thermo Fisher Scientific, Ontario, Canada

Introduction

Geographic atrophy

- Geographic atrophy (GA) is an advanced form of age-related macular degeneration.
- Approximately 1 million US adults are affected in at least one eye.[1]
- Risk increases with age (~0.3% of persons aged 65–74 years; ~4% of persons aged ≥ 85 years).[2]
- It is a progressive disease that results in irreversible loss of vision over time.[3]

Treatment

- Frequency of administration depends on the drug and regimen selected, but the current mainstay of treatment is intravitreal injections directly into the eye every 25–60 days.[4]
- The main goal of this treatment is to delay GA progression.
- While effective, studies have shown 30%–40% of patients discontinue treatment within a year.[5]

Problem

- As treatment is only effective if maintained, relatively high rates of discontinuation reduce patient long-term quality of life.[6]
- However, healthcare professionals do not typically have the time or resources to intensively monitor their patients, and thus support their patients' decisions to adhere to treatment.[5]
- Early discontinuation also affects the accuracy and reliability of clinical studies of treatment for GA.[7]
- Therefore, decision rules/algorithms targeting patient profiles at elevated risk of discontinuation may represent an efficient means to improve adherence and subsequently optimize positive outcomes.
- Specifically, if these algorithms were built on electronic, real-world, health data, they could be automated at various points-of-care to maximize their impact on treatment adherence.

Objectives

- **The main objective of this study was to develop a machine learning algorithm to identify patients at risk of dropout.**
- **For each patient, the algorithm would produce a risk score for adherence based on their individual patient journey.**
- **By utilizing a data-driven approach, we sought to explore the clinical, behavioral, and psychosocial factors behind patients' decisions to be non-adherent.**
- **The ultimate goal was to subsequently create a toolkit for healthcare professionals to implement a tailored retention strategy.**

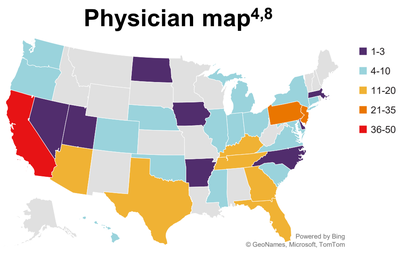
Methods

Data source – Vestrum retina database

Database overview[4,8]

Patient Visits (2015–Current)	18M
Patients (2015–Current)	2.4M
Sites (Current)	74
Retina Specialists (Current)	370
States (US) Represented (Current)	35

Physician map[4,8]



Patient follow-up time available[4,8]

Patients	2.4M
Total person years	4.1M
Mean follow-up time	1.7 Years

As of February 2024

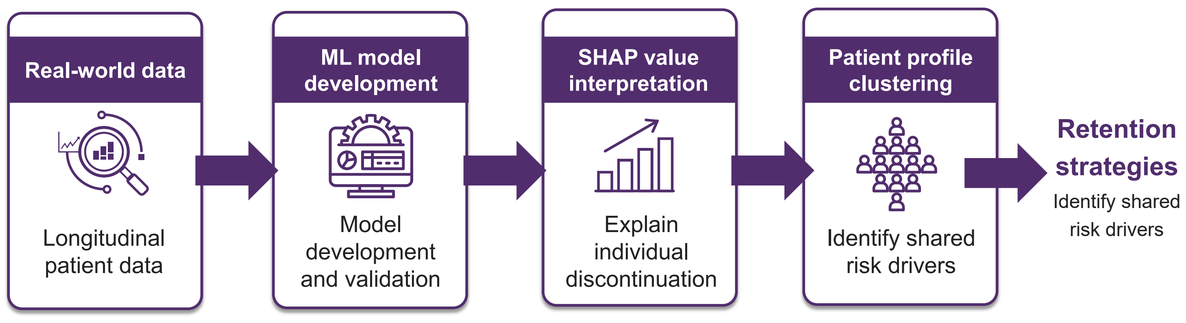
In addition, we can mine the free-text field within the electronic medical record (EMR) to discover further diagnostics and characteristics like subretinal fluid and vitreomacular traction (VMT)

- **Demographics:** Age, gender, smoking status, alcohol use
- Treatment history: Medications/treatment regimens (prior and present) and related procedures
- **Disease characteristics:** Relevant disease (e.g., GA) and comorbidities, year and severity of diagnosis, disease progression, adverse events
- **Patient-reported outcomes:** Visual acuity, central retinal thickness, fluid, surgery outcomes, change in therapy, lost to follow-up
- **Targeted adverse events:** infection (uveitis, endophthalmitis), retinal detachments, hemorrhaging, disease development, and progression

ML framework (Figure 1)

- We modelled early treatment discontinuation by training a supervised ML model based on 10,000+ GA patients; it was trained on 8,134 patients, tested on 2,034 patients (80–20% split)
- Model inputs included demographics, clinical history, and treatment patterns.
- We used SHapley Additive exPlanations (SHAP) to interpret patient-level risk drivers and applied K-means clustering on SHAP values to group similar patients. We identified five distinct patient profiles based on shared dropout risk factors.

Figure 1. Overview of risk profiling model and real-world impact



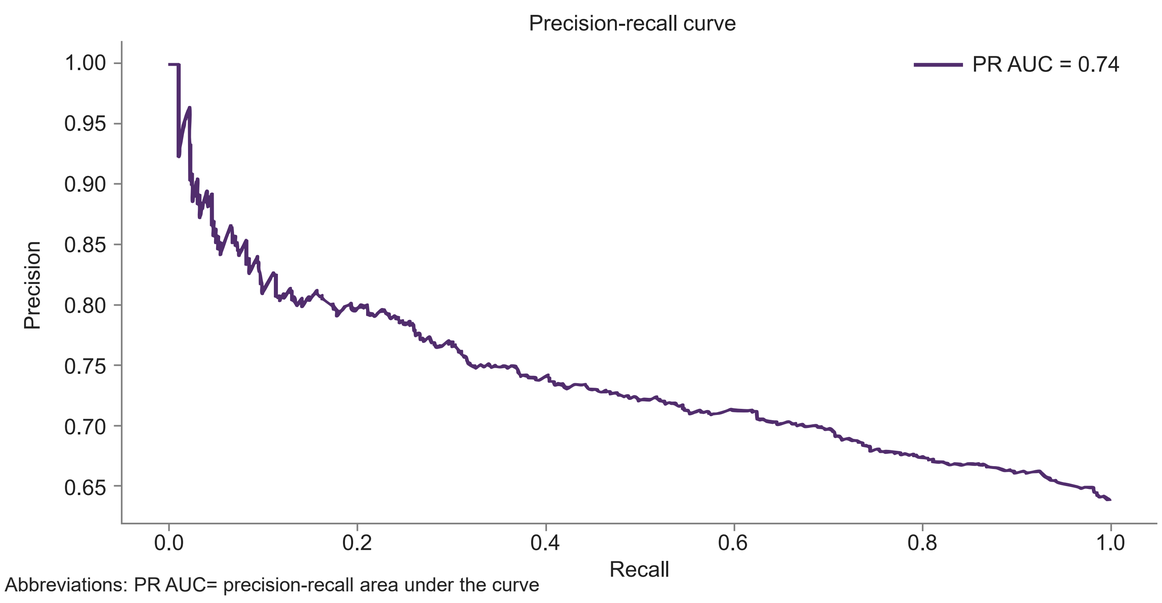
Abbreviations: ML = machine learning; SHAP = SHapley Additive exPlanations

Results

ML model results on patients at risk of GA treatment dropout (Figure 3)

- Prediction target was discontinuation within 120 days.
- The receiver operating characteristic (ROC) area under the curve (AUC) was 0.62; the precision-recall AUC was 0.74, with recall of 0.98 and precision of 0.64
- Overall accuracy was 64%.

Figure 3. Precision-recall curve by supervised machine learning model identifying patients at risk of GA treatment dropout



Abbreviations: PR AUC= precision-recall area under the curve

Patient profiles comparison (Figure 4)

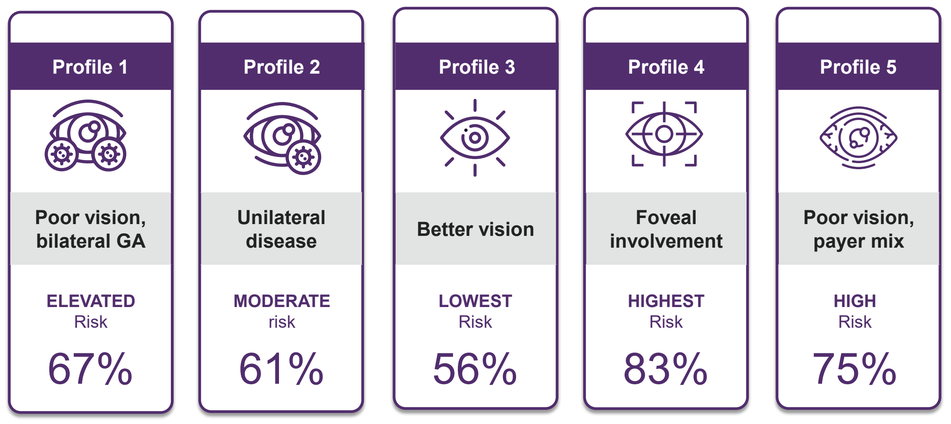
Profile 3—Lower GA treatment dropout

- 56% of patients in this cluster (which represents 26% of the total population) discontinued treatment within 120 days.
- They had relatively better initial vision, stable healthcare access, less severe disease (bilateral GA/fovea patterns), slower disease progression, and most were treatment naïve (i.e., first GA treatment ever).

Profile 5—Higher GA treatment dropout

- 75% of patients in this cluster (which represents 14% of the total population) discontinued treatment within 120 days.
- They had relatively poorer initial vision, limited/inadequate healthcare coverage, among other disease traits.

Figure 4. Patient profiles identified as at risk of GA treatment dropout



Conclusions

- **Patient profiles translated complex ML outputs and diverse real-world-data into clear, actionable subgroups.**
- **The profiles revealed both who is at risk or has a specific need, and why, enabling targeted strategies.**
- **The GA discontinuation example shows how prediction → explanation → segmentation can improve persistence and outcomes.**
- **Profiling is flexible—applicable to risk, engagement, value, and other use cases across the patient journey.**
- **Embedding profiles into workflows supports precision interventions, optimizes resources, and strengthens real-world evidence.**

References

1. Rein DB, et al. *JAMA Ophthalmol.* 2022;140(12):1202-8.; **2.** Vaz F, Picoto M. Geographic atrophy. 2017. https://amdbook.org/content/geographic-atrophy-0; **3.** Rajanala K, et al. *Front Ophthalmol (Lausanne).* 2023;3:1327883.; **4.** PPD™ clinical research business of Thermo Fisher Scientific. PPD Vestrum Health™ Real-World Data. https://www.ppd.com/our-solutions/clinical/real-world-data/vestrum-health-real-world-data/; **5.** Shahzad H, et al. *Syst Rev.* 2023;12(1):92.; **6.** Nissen AHK, et al. *Ophthalmol Ther.* 2025;14(6):1173-81.; **7.** Rosenblatt TR, et al. *Graefes Arch Clin Exp Ophthalmol.* 2022;260(1):93-100.; **8.** American Society of Retina Specialists. Retina image bank. https://imagebank.asrs.org/

Disclosures

All authors are employees of PPD™ Evidera™ Real-World Data & Scientific Solutions, Thermo Fisher Scientific. This poster was funded by Thermo Fisher Scientific.

Acknowledgment

Editorial and graphic design support were provided by Caroline Cole, Karissa Calara, and Richard Leason of Thermo Fisher Scientific.