## Verana Health<sub>®</sub>

Imputation of Disease Severity for Diabetic Retinopathy Clinical Encounters in the IRIS® Registry (Intelligent Research in Sight)

Richard Chapman, MS<sup>1</sup>, Thai Hien Nguyen, MPH<sup>1</sup>, Meghan Hatfield, MPH<sup>1</sup>, Theodore Leng, MD, MS<sup>1,2</sup>, Michael Mbagwu, MD<sup>1,2</sup>, Aracelis Z. Torres, PhD, MPH<sup>1</sup>, Durga S. Borkar, MD, MMCi<sup>1,3</sup>

<sup>1</sup>Verana Health, San Francisco, CA; <sup>2</sup>Byers Eye Institute, Stanford University School of Medicine, Palo Alto, CA; <sup>3</sup>Duke Eye Center, Duke University School of Medicine, Durham, NC



**Contact Information:** richard.chapman@veranahealth.com

#### Funding

This study was funded by Verana Health and all authors are employees or consultants of Verana Health.

## Purpose

- cause of irreversible blindness in the working age population.
- commonly leverage ICD codes to determine disease severity.
- coded for unspecified disease severity.
- performance of this model.

## Methods

- severity from 1/1/2014 to 9/30/2015 were included.
- prior clinical encounter, including known disease severity.
- DR (NPDR), or proliferative DR (PDR)

## Results

- clinical encounters were used.
- 0.950 in the test set.

## Conclusions

- disease severity in EHR clinical encounters.
- DR research.
- severity, appears to yield strong predictive power.

## • Diabetic retinopathy (DR) is a chronic, progressive disease and a leading

• Observational studies of DR using electronic health record (EHR) data

• Our prior work suggests a significant portion of EHR clinical encounters are

• The purpose of this study was to develop a model to impute DR disease severity using structured data fields in the EHR and to assess the

## • Patient encounters in the American Academy of Ophthalmology IRIS<sup>®</sup> Registry (Intelligent Research in Sight) with an ICD-9 code for DR with a known disease

• The dataset was randomized 80%/20% into training and test sets, respectively.

• A multinomial logistic regression model was developed utilizing structured demographic, clinical, and procedure data associated with the most recent

• DR disease severity was imputed as mild, moderate, severe nonproliferative

• Analyses were performed using Databricks (San Francisco, CA).

• Overall, 747,279 DR clinical encounters of 174,336 patients were included.

• A training set of 572,943 clinical encounters and a test set of 152,584

• The model achieved an accuracy of 96.9% and an aggregate F-1 score of

• These findings suggest that it may be possible to impute missing DR

• The ability to include all clinical encounters for a patient, including those with missing disease severity, could improve the quality of longitudinal

• Utilizing data from prior encounters, including those with known disease

# 600,000 400,000 200,000



## References

1. Klein BE. Overview of epidemiologic studies of diabetic retinopathy. Ophthalmic Epidemiol. 2007;14(4):179-183.

2. Lau M, Prenner JL, Brucker AJ, VanderBeek BL. Accuracy of Billing Codes Used in the Therapeutic Care of Diabetic Retinopathy. JAMA Ophthalmol. 2017;135(7):791-794.

3. Cai CX, Michalak SM, Stinnett SS, Muir KW, Fekrat S, Borkar DS. Effect of ICD-9 to ICD-10 Transition on Accuracy of Codes for Stage of Diabetic Retinopathy and Related Complications: Results from the CODER Study. Ophthalmol Retina. 2021;5(4):374-380.

4. Hatfield M, Nguyen TH, Chapman R, Myrick AC, Leng TL, Mbagwu M, Baxi S, Torres AZ, Borkar DS. Identifying the Mechanism of Missingness for Unspecified Diabetic Retinopathy Disease Severity in the Electronic Health Record: An IRIS Registry Analysis. J Am Med Inform Assoc. Ophthalmol. 2023 Mar 16. Online ahead of print.



### FIGURE 1: TRAINING AND TEST ENCOUNTERS BY SEVERITY

## TABLE 1: CONFUSION MATRIX

	ACTUAL DISEASE SEVERITY				
	MILD NPDR	MODERATE NPDR	SEVERE NPDR	PDR	PRECISION
ILD PDR	20,920	741	101	493	0.940
ATE PDR	511	19,190	303	615	0.931
ERE PDR	34	161	8,537	300	0.945
PDR	417	607	302	89,352	0.985
ALL	0.956	0.927	0.924	0.984	